

OM protein - protein search, using sw model

Run on: June 1, 2004, 14:32:13 ; Search time 30 seconds
(without alignments)
846.666 Million cell updates/sec

Title: US-09-615-285B-2

Perfect score: 2717
Sequence: 1 MALNSGSPPAIGPYENHGY.....YGNVMVFTDIYVQMRADG 492

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2717	100.0	492	3	US-09-342-749-2
2	2717	100.0	492	4	US-09-691-840-2
3	2696	99.2	492	4	US-09-685-166A-895
4	1540	56.7	283	3	US-08-807-151-1
5	1540	56.7	283	4	US-09-478-957-1
6	1165	42.9	209	4	US-09-685-166A-897
7	888	32.7	454	3	US-09-518-046-2
8	886	31.9	159	3	US-09-518-046-24
9	780.5	28.7	455	3	US-09-261-416-2
10	694	25.2	423	4	US-09-656-002-2
11	676.5	24.9	406	4	US-09-851-588-6
12	676.5	24.9	435	3	US-09-008-271A-6
13	676.5	24.9	437	4	US-09-851-588-8
14	660	24.3	798	1	US-08-200-900A-2
15	660	24.3	798	5	PCT-US94-00616-2
16	565.5	20.8	417	4	US-08-820-002-4
17	558.5	20.6	855	2	US-09-027-337-2
18	558.5	20.6	855	4	US-09-644-600-2
19	558.5	20.6	855	4	US-09-654-600A-2
20	557	20.5	376	4	US-09-820-002-2
21	556	20.5	638	2	US-08-681-151-3
22	553.5	20.4	248	3	US-08-944-483-63
23	548.5	20.2	902	4	US-09-644-600-10
24	548.5	20.2	902	4	US-09-654-600A-10
25	539.5	19.9	416	2	US-08-000-846-2
26	534	19.7	356	1	US-08-681-151-1
27	528.5	19.5	256	2	US-09-027-337-3

28	528.5	19.5	256	4	US-09-644-600-3	Sequence 3, Appli
29	528.5	19.5	256	4	US-09-654-600A-3	Sequence 3, Appli
30	527.5	19.4	418	1	US-08-508-448C-25	Sequence 25, Appl
31	527.5	19.4	418	4	US-09-370-838-62	Sequence 62, Appl
32	527.5	19.4	418	4	US-09-370-838-82	Sequence 82, Appl
33	527.5	19.4	418	4	US-09-370-838-83	Sequence 83, Appl
34	524	19.3	238	3	US-08-944-483-64	Sequence 64, Appl
35	523.5	19.3	255	3	US-08-944-483-67	Sequence 67, Appl
36	521.5	19.2	235	3	US-08-807-151-3	Sequence 3, Appli
37	521.5	19.2	235	4	US-09-478-357-3	Sequence 3, Appli
38	518.5	19.1	812	1	US-08-248-629A-1	Sequence 1, Appli
39	518.5	19.1	812	1	US-08-451-932-1	Sequence 1, Appli
40	518.5	19.1	812	1	US-08-452-360-1	Sequence 1, Appli
41	518.5	19.1	812	1	US-08-326-785-1	Sequence 1, Appli
42	518.5	19.1	812	2	US-08-612-788-1	Sequence 1, Appli
43	518.5	19.1	812	2	US-08-605-598B-1	Sequence 1, Appli
44	518.5	19.1	812	2	US-08-429-743-1	Sequence 1, Appli
45	518.5	19.1	812	2	US-08-866-735-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-342-749-2
; Sequence 2, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: Tumor suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-749-2

Query Match	100.0%;	Score 2717;	DB 3;	Length 492;
Best Local Similarity	100.0%;	Pred. No. 5.3e-244;		
Matches 492;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MALNSGSPPAIGPYENHGYQENPYPAQPTVVPTVYVHPAQYVPSVPQYAPRVLTOA	60	
DB	1	MALNSGSPPAIGPYENHGYQENPYPAQPTVVPTVYVHPAQYVPSVPQYAPRVLTOA	60	
QY	61	SNPVCTQPKSPSGTCTKTKKALCITLTGLFTLVGAALAGLLWKFMGSKCSNGIEC	120	
DB	61	SNPVCTQPKSPSGTCTKTKKALCITLTGLFTLVGAALAGLLWKFMGSKCSNGIEC	120	
QY	121	DSGTCINPSNWCDSVSHCPGGEDENRCVRLYGPNFILQVYSSQKSWHPVQDDWNNY	180	
DB	121	DSGTCINPSNWCDSVSHCPGGEDENRCVRLYGPNFILQVYSSQKSWHPVQDDWNNY	180	
QY	181	GRAACDMGYKNFYSSQGIIVDSGTSFMKLNLSAGNVDIYKLYHSDACSSKAVWSLR	240	
DB	181	GRAACDMGYKNFYSSQGIIVDSGTSFMKLNLSAGNVDIYKLYHSDACSSKAVWSLR	240	
QY	241	CIACGVNLNSRSRQIRVGGESALPGAWPQVSLHVNHVHVCSSIIITPEWITTAHCVEK	300	
DB	241	CIACGVNLNSRSRQIRVGGESALPGAWPQVSLHVNHVHVCSSIIITPEWITTAHCVEK	300	
QY	301	PLANPHTWTAFAQILQSFYAGVQVEKVTSHPNYDSKTKNDIALMKLQKPLTFNDL	360	
DB	301	PLANPHTWTAFAQILQSFYAGVQVEKVTSHPNYDSKTKNDIALMKLQKPLTFNDL	360	

QY 361 VKPVCLPFGMLQPEQLCWISGKATEEKGKTSVNAKVLIIETQRCNSRYVDNLI 420
DB 361 VAPVCLPFGMLQPEQLCWISGKATEEKGKTSVNAKVLIIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLOGNVDSQCGSGPLVTSKNNIWLIGDTSWGGCAKAYRPGVYGNVMVF 480
DB 421 TPAMICAGFLOGNVDSQCGSGPLVTSKNNIWLIGDTSWGGCAKAYRPGVYGNVMVF 480
QY 481 TDWIYQNRADG 492
DB 481 TDWIYQNRADG 492

RESULT 2
US-09-691-840-2
; Sequence 2, Application US/09691840
; Patent No. 644419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2118-202
; CURRENT APPLICATION NUMBER: US/09/691,840
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342,749
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091,044
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-691-840-2

Query Match 100.0%; Score 2717; DB 4; Length 492;
Best Local Similarity 100.0%; Pred. No. 5.3e-244;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNSGPPAIGPYENHGYQENPYPAQPTVTVYVHVPAYQYPSVPQVAPRVLTOA 60
DB 1 MALNSGPPAIGPYENHGYQENPYPAQPTVTVYVHVPAYQYPSVPQVAPRVLTOA 60
QY 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAGLLWKFGKCSNSGIEC 120
DB 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAGLLWKFGKCSNSGIEC 120
QY 121 DSSGTCINPNCWCDGVSHCPGGEDENRVLYGPNFLOVYSSQKSHHPVCQDDWNNY 180
DB 121 DSSGTCINPNCWCDGVSHCPGGEDENRVLYGPNFLOVYSSQKSHHPVCQDDWNNY 180
QY 181 GRAACRDMGYKNFYSSQGIYVDDSGSTSPMKLNTSAGNVDIYKLYHSDACSCKAVVSLR 240
DB 181 GRAACRDMGYKNFYSSQGIYVDDSGSTSPMKLNTSAGNVDIYKLYHSDACSCKAVVSLR 240
QY 241 CIACGVNLNSRQSRIRVGGESALPGAWPQVSLHVNQVHVCVGSIIIPWIVTAACHVEK 300
DB 241 CIACGVNLNSRQSRIRVGGESALPGAWPQVSLHVNQVHVCVGSIIIPWIVTAACHVEK 300
QY 301 PLANNPMTAFAGILRQSFMYGAGYQVEKVIHSHPNYDSTKNNDIAMKLOKPLTFNDL 360
DB 301 PLANNPMTAFAGILRQSFMYGAGYQVEKVIHSHPNYDSTKNNDIAMKLOKPLTFNDL 360
QY 361 VKPVCLPFGMLQPEQLCWISGKATEEKGKTSVNAKVLIIETQRCNSRYVDNLI 420
DB 361 VKPVCLPFGMLQPEQLCWISGKATEEKGKTSVNAKVLIIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLOGNVDSQCGSGPLVTSKNNIWLIGDTSWGGCAKAYRPGVYGNVMVF 480

DB 421 TPAMICAGFLOGNVDSQCGSGPLVTSKNNIWLIGDTSWGGCAKAYRPGVYGNVMVF 480
QY 481 TDWIYQNRADG 492
DB 481 TDWIYQNRADG 492

RESULT 3
US-09-685-166A-895
; Sequence 895, Application US/09685166A
; Patent No. 8630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C21
; CURRENT APPLICATION NUMBER: US/09/685,166A
; CURRENT FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-685-166A-895

Query Match 99.2%; Score 2696; DB 4; Length 492;
Best Local Similarity 98.8%; Pred. No. 4.7e-242;
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALNSGPPAIGPYENHGYQENPYPAQPTVTVYVHVPAYQYPSVPQVAPRVLTOA 60
DB 1 MALNSGPPAIGPYENHGYQENPYPAQPTVTVYVHVPAYQYPSVPQVAPRVLTOA 60
QY 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAGLLWKFGKCSNSGIEC 120
DB 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAGLLWKFGKCSNSGIEC 120
QY 121 DSSGTCINPNCWCDGVSHCPGGEDENRVLYGPNFLOVYSSQKSHHPVCQDDWNNY 180
DB 121 DSSGTCINPNCWCDGVSHCPGGEDENRVLYGPNFLOVYSSQKSHHPVCQDDWNNY 180
QY 181 GRAACRDMGYKNFYSSQGIYVDDSGSTSPMKLNTSAGNVDIYKLYHSDACSCKAVVSLR 240
DB 181 GRAACRDMGYKNFYSSQGIYVDDSGSTSPMKLNTSAGNVDIYKLYHSDACSCKAVVSLR 240
QY 241 CIACGVNLNSRQSRIRVGGESALPGAWPQVSLHVNQVHVCVGSIIIPWIVTAACHVEK 300
DB 241 CIACGVNLNSRQSRIRVGGESALPGAWPQVSLHVNQVHVCVGSIIIPWIVTAACHVEK 300
QY 301 PLANNPMTAFAGILRQSFMYGAGYQVEKVIHSHPNYDSTKNNDIAMKLOKPLTFNDL 360
DB 301 PLANNPMTAFAGILRQSFMYGAGYQVEKVIHSHPNYDSTKNNDIAMKLOKPLTFNDL 360
QY 361 VKPVCLPFGMLQPEQLCWISGKATEEKGKTSVNAKVLIIETQRCNSRYVDNLI 420
DB 361 VKPVCLPFGMLQPEQLCWISGKATEEKGKTSVNAKVLIIETQRCNSRYVDNLI 420

QY 421 TPAMICAGFLQGNVDSQDGGPLVTSKNNIWMWLGDTSMGSCAKAYRQVYGNVWF 480
Db 421 TPAMICAGFLQGNVDSQDGGPLVTSKNNIWMWLGDTSMGSCAKAYRQVYGNVWF 480
QY 481 TDWIYQMRADG 492
Db 481 TDWIYQMRANG 492

RESULT 4

US-08-807-151-1
; Sequence 1, Application US/08807151
; Patent No. 6043033
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,151
; FILING DATE: Filed Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNOT01
; CLONE: 556016
US-08-807-151-1

Query Match 56.7%; Score 1540; DB 3; Length 283;
Best Local Similarity 99.6%; Pred. No. 6.1e-135;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 210 MKLNTSAGNVDIYKKLYHSDACSSKAVVSLRCLACGVNLSRSQSRIVGGESALPGAWPW 269
Db 1 MKLNTSAGNVDIYKKLYHSDACSSKAVVSLRCLACGVNLSRSQSRIVGGESALPGAWPW 60
QY 270 QVSLHVQNVHVCVGGSIITPEWIVTAAHCVKELNPNPWHWTAFAGILRQSFMYGAGYQVE 329
Db 61 QVSLHVQNVHVCVGGSIITPEWIVTAAHCVKELNPNPWHWTAFAGILRQSFMYGAGYQVE 120
QY 330 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLNPGMMLQPEQLCWISGWGATEE 389
Db 121 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLNPGMMLQPEQLCWISGWGATEE 180

QY 390 KGKTSVLNAAKVLLIETORCNSRVYVDNLITPAMICAGFLQGNVDSQDGGPLVTSK 449
Db 181 KGKTSVLNAAKVLLIETORCNSRVYVDNLITPAMICAGFLQGNVDSQDGGPLVTSK 240
QY 450 NNIMWLGDTSMGSCAKAYRQVYGNVWVFTDIYQMRADG 492
Db 241 NNIMWLGDTSMGSCAKAYRQVYGNVWVFTDIYQMRADG 283

RESULT 5

US-09-478-957-1
; Sequence 1, Application US/09478957
; Patent No. 6350448
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/478,957
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/807,151
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNOT01
; CLONE: 556016
US-09-478-957-1

Query Match 56.7%; Score 1540; DB 4; Length 283;
Best Local Similarity 99.6%; Pred. No. 6.1e-135;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 210 MKLNTSAGNVDIYKKLYHSDACSSKAVVSLRCLACGVNLSRSQSRIVGGESALPGAWPW 269
Db 1 MKLNTSAGNVDIYKKLYHSDACSSKAVVSLRCLACGVNLSRSQSRIVGGESALPGAWPW 60
QY 270 QVSLHVQNVHVCVGGSIITPEWIVTAAHCVKELNPNPWHWTAFAGILRQSFMYGAGYQVE 329
Db 61 QVSLHVQNVHVCVGGSIITPEWIVTAAHCVKELNPNPWHWTAFAGILRQSFMYGAGYQVE 120
QY 330 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLNPGMMLQPEQLCWISGWGATEE 389
Db 121 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLNPGMMLQPEQLCWISGWGATEE 180
QY 390 KGKTSVLNAAKVLLIETORCNSRVYVDNLITPAMICAGFLQGNVDSQDGGPLVTSK 449

Db 181 KGKTSVLNAKVLLETORCNSRVVYDNLITPAMICAGFLQGNVDSQGSGLVTSK 240
Qy 450 NNIMWLIGDTSWGCACAYRPGVYGNVMTDVIYROMRADG 492
Db 241 NNIMWLIGDTSWGCACAYRPGVYGNVMTDVIYROMRADG 283

RESULT 6

US-09-685-166A-897
; Sequence 897, Application US/09685166A
; Patent No. 6630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C21
; CURRENT APPLICATION NUMBER: US/09/685,166A
; NUMBER OF SEQ ID NOS: 10-10
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 897
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-685-166A-897

Query Match 42.9%; Score 1165; DB 4; Length 209;
Best Local Similarity 99.5%; Pred. No. 2.8e-100; Indels 0; Gaps 0;
Matches 208; Conservative 1; Mismatches 0

Qy 1 MALNSGPPAIGPYENHGYQENPYPAQPTVPTVYVHPAQYPPVQYAPRVLTOA 60
Db 1 MALNSGPPAIGPYENHGYQENPYPAQPTVPTVYVHPAQYPPVQYAPRVLTOA 60
Qy 61 SNPVTCTPKSPGVTCTSKKALCITLTGTLVGAALAGLLKFMGSKNSGIEC 120
Db 61 SNPVTCTPKSPGVTCTSKKALCITLTGTLVGAALAGLLKFMGSKNSGIEC 120
Qy 121 DSSGTCINPSNWCDSVHCPGSDENRCVRLYGNPFIQVYSSQKSWHPVQDDWNNY 180
Db 121 DSSGTCINPSNWCDSVHCPGSDENRCVRLYGNPFIQVYSSQKSWHPVQDDWNNY 180
Qy 181 GRAACRDMGYKNFYSSGIVDDSGSTSF 209
Db 181 GRAACRDMGYKNFYSSGIVDDSGSTSF 209

RESULT 7

US-09-518-046-2
; Sequence 2, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof

FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 2
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: complete amino acid sequence of TADG-12
; OTHER INFORMATION: protein
US-09-518-046-2

Query Match 32.7%; Score 888; DB 3; Length 454;
Best Local Similarity 46.3%; Pred. No. 4.9e-74;
Matches 190; Conservative 52; Mismatches 148; Indels 20; Gaps 10;

Qy 90 TLGTFVGAALAGLLKFMGSKNSGIECDSGTCINPSNWCDSVHCPGSDENRCV 149
Db 54 SLGIILALALAGLGIHF---DCSGK-YRCRSFKCIELITRCDGVSDCKDGEVRCV 109
Qy 150 RLYGFNFILQVYSSQKSWHPVQDDWNNYGRAACRDMGYKNFYSSQKSWHPVQDDSGSTSF 209
Db 110 RVGGONAVLQVFTA--ASWKTMSDDWKGHYANVACAQLGFP-SYVSSDNLRYVSSLEGQF 166
Qy 210 MKLNTSAGNV---DIYKLYHS---DACSSKAVSVSLRCIACGVNLNSRQSRIVGESA 262
Db 167 REEFVSDIDLLPDDKVTALHSHVYREGCASGHVILQCTACGHRGYS--SRIVGNMS 224
Qy 263 LPGAHPQVSLHVNQVHVCVGSITPEWIVTAACHVEKPLNNPWHMTAFAGILRQSFMY 322
Db 225 LLSQWPMQASLQFGYHLCGSSVITPLWITIAAHCV-YDLYLPKSWTIQVGLV--SILLDN 281
Qy 323 GA-GYQVEKIVSHENYDSKTKNDIALMKLQKLPILTENDLVKPYCLNPGMWLQPEQLCWI 381
Db 282 PAPSHLVEKIVYHSKYKPKRLGNDIALMKLAGLPTFNEMIQPVCLNSEENFPDGVKVCWT 341
Qy 382 SGWGATEEKGTSEVLNAAKVLLIETORCNSRVYVYDNLITPAMICAGFLQGNVDSQGS 441
Db 342 SGWGATEDGGDASPLNAHAAPLISNKNHRDVGGIISPSMLCAGYLTGGVDSQGS 401
Qy 442 GGPLVTSKNIMWLIGDTSWGCACAYRPGVYGNVMTDVIYROMRAD 491
Db 402 GGPLVTCERRLWKLVGATSFQIGCAEVNKGIVTRVTSFLDNIHQMERD 451

RESULT 8

US-09-518-046-24
; Sequence 24, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 24
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; OTHER INFORMATION: protease domain of TMPRSS2 (Tmprs2)
US-09-518-046-24

Query Match 31.9%; Score 866; DB 3; Length 159;

Best Local Similarity 99.4%; Pred. No. 1.2e-72;
Matches 158; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 289 ENIVTAACHVCKPLNNPWHWTAFAGILRQSFMYGAGYQVEKVISHPNVDSTKNDIAL 348
Db 1 ENIVTAACHVCKPLNNPWHWTAFAGILRQSFMYGAGYQVEKVISHPNVDSTKNDIAL 60
QY 349 MKLQKPLFNDLVKPVCLPNNPMMQLPEQLCWISGAGTEKGTSEVLNAKVLIIETQ 408
Db 61 MKLQKPLFNDLVKPVCLPNNPMMQLPEQLCWISGAGTEKGTSEVLNAKVLIIETQ 120
QY 409 RNSRYVVDNLITPAMICAGFLOGNVDSGCGSGPLVT 447
Db 121 RNSRYVVDNLITPAMICAGFLOGNVDSGCGSGPLVT 159
RESULT 9
US-09-261-416-2
; Sequence 2, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TAGD-12: A No. 6291663el Transmembrane Serine Protease
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261.416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 2
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TAGD-12 encoded by nucleotides
; OTHER INFORMATION: 144 to 1511 of Sequence 1
; Patent No. 6291663
US-09-261-416-2
Query Match 25.7%; Score 780.5; DB 3; Length 455;
Best Local Similarity 43.6%; Pred. No. 4.8e-64;
Matches 180; Conservative 50; Mismatches 158; Indels 25; Gaps 13;
QY 90 TLGFLVGAALAAAGLLMKFGKSCNSGIECDSSGTCINPNSWCDGVSHCPGGEDEKCV 149
Db 54 SLGIIALLALAIGLGTHF--DCSGK-YRCRSFKCIBLITRCDGVSDCKDGEDEYRCV 109
QY 150 RLYGFNFILQYSSQRKSHFVCCDDNENYGRAACRDMGYKXNFYSSQGIYDDSGSTF 209
Db 110 RVGGQNAVLOVFTA--ASWTKTMCDDWKGHYANVACAQLGFP--SYVSSDNLNVRVSLGQF 166
QY 210 MKLNTSAGNV---DIYKLYHS---DACS KAVVSLRITACGVNLNRSRQRIYGGESA 262
Db 167 REEFVSDHLLPDDKVTAHSHVYVREGCASGHVVTLOCTACGHRGYS--SRIYVGNWS 224
QY 263 LPGAWPQVSLHVQNVHVCSSGIIITPEWIVTAACHVCKEPLNPNWHWTAFAGILRQSFMY 322
Db 225 LLSQMPWQASLQFGYHLCGGSVITPLWITAAHCY-YDLYLPKNTIQVGLV--SLLDN 281
QY 323 GA-GYQVEKVISHPNVDSTKNDIALMKLQKPLFNDLVKPVCLPNNPMMQLPEQLCWI 381
Db 282 PAPSHLVKIVYHSKYKPKELGNDIALMKLAGELTNEMIQVCLPNSSEENPPDKVCWT 341
QY 382 SGWAGTEKGTSEVLNAKVLIIETQ--RNSRYVVDNLITPAMICAGFLOG-NVDSQ 438
Db 342 SGWAGTEGDDGASPLVNAHAAPVLSINKDLPQQRVWRHHL--PLHALRGLPDGWRWNSQ 399
QY 439 GDSGGLVTSKNNIWLIGTSGWGCACAYRPGVYGNVWFTDIYQMRAD 491
Db 400 GDSGGLVQCERRLKLVGATSGIGCADVNEKGVVTRVTSFLDNIHQMERD 452

RESULT 10

US-09-656-002-2
; Sequence 2, Application US/09656002
; Patent No. 6455668
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF SCREENING FOR COLORECTAL CANCER, COMPOSITIONS, AND
; FILE REFERENCE: A-69108/DJ8/JUD/AMS
; CURRENT APPLICATION NUMBER: US/09/656,002
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/525,993
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US 09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US 00/07044
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-656-002-2
Query Match 25.2%; Score 684; DB 4; Length 423;
Best Local Similarity 36.2%; Pred. No. 4e-55;
Matches 165; Conservative 64; Mismatches 167; Indels 60; Gaps 16;
QY 61 SNPVCCTPKSPGTVCTKTKKALCITLTGLTFLVGAALAAAGLLMKFGKSCNSGIEC 120
Db 2 SNP--CANPVSP-WRPSESVGIPLIALLSLASIIIVVLLIKVILDKYY-----FLC 50
QY 121 DSSGTCINPNSWCDGVSHCPGGEDEKCVRLY--GP-----NFTLOVYSSORSHW 169
Db 51 GQPLHFIPKQCLDGLDCLPGLGEDEHCVKSPFPGPAVAVLRSKDRSTLOVLDATGNWF 110
QY 170 PVCQDDNENYGRAACRDMGY--KXNFYSSQ-----GIYDDSGSTSFMKLNTSAGNYD 220
Db 111 SADFONFTALAEATACRQWYSSKPTFRAVEIGPDQLDVVEITENSQELMRNMSG-- 167
QY 221 IYKLYHS DACSKAVVSLRITACGVNLNRSRQRIYGGESALPGAWPQVSLHVQNVH 280
Db 168 -----PCLSGSGLVSLHCLACGSL---KTRFVVGGEASVDSHPWQVSIQYDKQHV 215
QY 281 CGSIIITPEWIVTAACHVCKEPLNPNWHWTAFAGILR-QSFMYGAGYQVEKVI---SHPN 336
Db 216 CGSILDPHVLTAACHCFRKH-TDVFNNKVRAGSDKLGSP-----PSLAVAKIIIEFNP 270
QY 337 YDSKTKNNDIALMKLQKPLFNDLVKPVCLPNNPMMQLPEQLCWISGAGTEK-GKTSE 395
Db 271 Y---PKNDIALMKLQKPLFSTGTVRPICLFPDEELTPATPLWICGWFTKQNGKMSD 327
QY 396 VLNAKVLIIETQRCNSRYVVDNLITPAMICAGFLOGNVDSGCGSGPLVTSKNNIWL 455
Db 328 ILLQASQVQIDSTRCNADDAVCGEVTCKMCCAGPEGGVDTCCGDSGGLMYOSDQ-WHV 386
QY 456 IGDTSGWGCACAYRPGVYGNVWFTDIYQMRAD 491
Db 387 VGIVSWGYGCGSPSTFGVYTKVSAYLNNIYNWKA 422
RESULT 11
US-09-851-588-6
; Sequence 6, Application US/09851588
; Patent No. 6682890
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF SCREENING FOR COLORECTAL CANCER, COMPOSITIONS, AND
; FILE REFERENCE: A-69108/DJ8/JUD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/525,993
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US 09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US 00/07044
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-851-588-6

FILE REFERENCE: A-68829-1/DJB/JJD/AMS
 CURRENT APPLICATION NUMBER: US/09/851,588
 FILING DATE: 2001-09-24
 PRIOR APPLICATION NUMBER: US 09/642,252
 PRIOR FILING DATE: 2000-08-17
 PRIOR APPLICATION NUMBER: US 09/656,002
 PRIOR FILING DATE: 2000-09-06
 NUMBER OF SEQ ID NOS: 9
 SOFTWARE: Patent in version 3.1
 SEQ ID NO 6
 LENGTH: 406
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-851-588-6

Query Match 24.9%; Score 676.5; DB 4; Length 406;
 Best Local Similarity 39.1%; Pred. No. 1.9e-54;
 Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;
 QY 133 CDGVSHCPGGDENRCVRLY--GP-----NFILQVSSQKSKHPVQCDDWNYG 181
 DB 46 CDGELDCPLGDEEHCKVSPFPGPAVAVRLSKDRSTLQVLSAYGNWFSACFDNTEALA 105
 QY 182 RAACRDNGY--KXNFYSSQ-----GIVDDSGSTFMKLNSTAGNVDIYKLYHSDACS 232
 DB 106 ETACRQMGYSKPTTFAVEIGDPDQDLVDVVEITENSQELRMNSG-----PCL 153
 QY 233 SKAVVSLRCIACGVNLNSRQSRIVGGESALPGAMPQVSLHVQNVHVCSSIIITPEWIV 292
 DB 154 SGLSVSLHCLACKSL---KTRPVVGGEASVDSWPMQVSIQYDKQVCGSILDPHWL 210
 QY 293 TAAHCVKPLNPNHWTAFAGILR--QSFMYGAGYQVEKVI---SHPNYDSKTKNDIAL 348
 DB 211 TAAHCFRKH--TDVFNWVKRAGSKLGSF---PSLAVAKIIIEFNPMY---PRNDIAL 262
 QY 349 MKLQKPLTFNDLVKPVCLPNPMMLOPEQLCWISGNGATEEK-GKTSEVLNAAKVLIIET 407
 DB 263 MKLQPLTFSGTVRPICLPFPDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDS 322
 QY 408 QRCNSRYVDNLITPAMICAGFLOGNVDSQCGSGGLPVTSKNIWMLIGDTSWGSCAK 467
 DB 323 TRCNADDAIYQGEVTEKMMKAGIPGGVDTCCGDSGGPLMYQSDQ--WHVVGIVSWGCGG 381
 QY 468 AYRPGVYGNVMTDIYRQMRAD 491
 DB 382 PSTPGVYTKVSAYLNWYINWKA 405

RESULT 12

US-09-008-271A-6
 Sequence 6, Application US/09008271A
 Patent No. 6203979
 GENERAL INFORMATION:

APPLICANT: Bandman, Olga
 Hillman, Jennifer L.
 Yue, Henry
 Guegler, Karl J.
 Corley, Neil C.
 Tang, Tom Y.
 Shah, Purvi

TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/008,271A
 FILING DATE: 16-Jan-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: <Unknown>
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Mohan-Peterson, Sheila
 REGISTRATION NUMBER: 41,201
 REFERENCE/DOCKET NUMBER: PF-0458 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-855-0555
 TELEFAX: 650-845-4166
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 435 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: COLN0T13
 CLONE: 1337018
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 US-09-008-271A-6

Query Match 24.9%; Score 676.5; DB 3; Length 435;

Best Local Similarity 39.1%; Pred. No. 2.1e-54;
 Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;

QY 133 CDGVSHCPGGDENRCVRLY--GP-----NFILQVSSQKSKHPVQCDDWNYG 181
 DB 75 CDGELDCPLGDEEHCKVSPFPGPAVAVRLSKDRSTLQVLSAYGNWFSACFDNTEALA 134
 QY 182 RAACRDNGY--KXNFYSSQ-----GIVDDSGSTFMKLNSTAGNVDIYKLYHSDACS 232
 DB 135 ETACRQMGYSKPTTFAVEIGDPDQDLVDVVEITENSQELRMNSG-----PCL 182
 QY 233 SKAVVSLRCIACGVNLNSRQSRIVGGESALPGAMPQVSLHVQNVHVCSSIIITPEWIV 292
 DB 183 SGLSVSLHCLACKSL---KTRPVVGGEASVDSWPMQVSIQYDKQVCGSILDPHWL 239
 QY 293 TAAHCVKPLNPNHWTAFAGILR--QSFMYGAGYQVEKVI---SHPNYDSKTKNDIAL 348
 DB 240 TAAHCFRKH--TDVFNWVKRAGSKLGSF---PSLAVAKIIIEFNPMY---PRNDIAL 291
 QY 349 MKLQKPLTFNDLVKPVCLPNPMMLOPEQLCWISGNGATEEK-GKTSEVLNAAKVLIIET 407
 DB 292 MKLQPLTFSGTVRPICLPFPDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDS 351
 QY 408 QRCNSRYVDNLITPAMICAGFLOGNVDSQCGSGGLPVTSKNIWMLIGDTSWGSCAK 467
 DB 352 TRCNADDAIYQGEVTEKMMKAGIPGGVDTCCGDSGGPLMYQSDQ--WHVVGIVSWGCGG 410
 QY 468 AYRPGVYGNVMTDIYRQMRAD 491
 DB 411 PSTPGVYTKVSAYLNWYINWKA 434

RESULT 13

US-09-851-588-8

Sequence 8, Application US/09851588

Patent No. 6682890

GENERAL INFORMATION:

APPLICANT: Mack, David

APPLICANT: Gish, Kurt C.

APPLICANT: Wilson, Keith E.

TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND

TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS

FILE REFERENCE: A-68829-1/DJB/JJD/AMS

CURRENT APPLICATION NUMBER: US/09/851,588

CURRENT FILING DATE: 2001-09-24

PRIOR APPLICATION NUMBER: US 09/642,252

; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 8
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-851-588-8

Query Match 24.9%; Score 676.5; DB 4; Length 437;
Best Local Similarity 39.1%; Pred. No. 2.1e-54;
Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;
QY 133 CDGVSCHPCGDEHRCVRLY--GP-----NFILQVSSQKSWHPVQCDDWNNY 181
DB 77 CDGELDCPLGDEHCHVKSPGPAVAVLSKDRSTLQVLDATGNCWFSAFCDFNTEALA 136
QY 182 RAACRMGY--KXNVSSQ-----GIVDSGTSFVKMLNTSAGNVDIYKLVHSACS 232
DB 137 ETACROMGYSSKPTTFAVIGIPDODLDVVEITENSQELMRNSSG-----PCL 184
QY 233 SKAVSLRCAQVNLNSRQRIYVGSALPGAMPQVSLHVQNVHVCSSIIITPWI 292
DB 185 SGLSLVHLCLAGKSL--KTRVVGGEASVDSWPAQVSIQYKQHVCGSILDPHWL 241
QY 293 TAACHVEKELNPNHWTAPAGILR--QSPMFYAGYQVEKVI---SHPNYDSKTKNDIAL 348
DB 242 TAACHCRPKH--TDVFNKVRAGSKLGSF---PSLAVAKIIIEFNPMY----PRNDIAL 293
QY 349 MKLQKLTENDLVKPCVCLNPGMMLQPEQLCWISQWATEEK-KGTSEVLNAAKVLLET 407
DB 294 MKLQFLTSGVTRPCLPFPDEELTPATPLWLGWTKQGGKMSDILLQASVQVIDS 353
QY 408 QRCNSRYVDNLITPAMICAGFLQNVDSQSGSGLPLVTSKNNIWMIGTSGSCAK 467
DB 354 TRCNADDAYCGEVTENKMCAGIPEGVDTCQDGGPLMYSDQ-WHVVGIVSWGCGG 412
QY 468 AVRPGVYGVNMFVDIYRQMDAD 491
DB 413 PSTPGVYTKVSAYLNWYNWKA 436

RESULT 14
US-08-200-900A-2
; Sequence 2, Application US/08200900A
; Patent No. 565556
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,900A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinerdt, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 876-1170 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-200-900A-2

Query Match 24.3%; Score 660; DB 1; Length 798;
Best Local Similarity 33.9%; Pred. No. 1.7e-52;
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;
QY 65 VCTQKPSGTCVCTSKKALCIITL-----GTELVGAALAAAGLLWKFMSKCSNSG 117
DB 374 VYTGPGVNDVFSITNRMTVLFITDMLAKQGFKANETTVGLG-----IPEPCKEDN 426
QY 118 IECDSSTGTCINPSKWCQGVSHCPGGEDEHRCVRLY-----PNFILQVSSQKSWH 169
DB 427 FQC-KDGEICPLVNLCDGFPCHCKDGSDEAHCVRLFNCTTDSGLVQFRIQ-----SIWH 479
QY 170 PVOQDDWNNYGRAACRDMGYKNNFYSSQGIIVDDSGTSPFKMLNTSAGNVDIYKLVHSD 229
DB 480 VACAENWTQISDDVCQLLGLGTG--NSSVPTSTGGPYVNLNTAPNGSLI---LTPSQ 534
QY 230 ACSSKAVVSLRC--IACQVNLNSRQS-RIVGSSALPGAMPQVSLHVQNVHVCSSII 286
DB 535 QCLEDSLILQCNYSKCGKLVTOEVSFKIVGSDSREGAMPWVVALYFDDQVCGASLV 594
QY 287 TPEWTVTAACHVCKPLNPNWETAFAGILRQSFYAGYQVE-----KVISHPNYDSKT 341
DB 595 SEDMLVSAACHVYGENMEPSKKAVALGLHMASNL---TSPQIETRLDQIVINPHYNKRR 651
QY 342 KNDIALMKLQKLTENDLVKPCVCLNPGMMLQPEQLCWISQWATEEKTEKTEVLNAAK 401
DB 652 KNDIAMHLEMKVNYDIYQICLPEENQVFPFGKICSIAGMGALYQSGADVLQEAD 711
QY 402 VLLIETORCNSRYVDNLITPAMICAGFLQNVDSQSGSGLPLVTSKNNIWMIGTSG 461
DB 712 VPLISNEKQQQOMPEYN--IENMVACAGYEAGGVDSQSGSGLPLVTSKNNIWMIGTSG 770
QY 462 SGGCAKAYRPGVYGVNMFVDI 484
DB 771 GYQCALPNRPGVYARVFRFTWI 793

RESULT 15
PCT-US94-00616-2
; Sequence 2, Application PC/TUS9400616
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 33
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00616
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-00616-2

Query Match 24.3%; Score 660; DB 5; Length 798;
Best Local Similarity 33.9%; Pred. No. 1.7e-52;
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;

Qy	65	VCTQPKSPSGTCTSKTKALCITLTL-----GTFLVGAALAGLLWKFMGSKCSNSG	117
Db	374	VYTGPGFVNDVFSITTNRMTVLFITDNMLAKQGPKANFTTGYGLG-----IPEPCKEDN	426
Qy	118	IBCDSSGTCINPNWCDGVSHCPGDEENRCVRLYG-----PNFILOVYSSQKSKWH	169
Db	427	FQC-KDGEICPLVNLCDGFFHCKDGSDEAHCVLFFNGTTDSSGLVQFRIQ-----SIWH	479
Qy	170	PVQDDWNNENYGRAACRDMGYKNFYSSQGI VDDSGTSFPMKLNTSAGNVDIYKKLYHSD	229
Db	480	VACAEWTTQISDDVCOLLGLGTG--NSSVPTSTGGPGPVNLTAPNGSLI---LTPSQ	534
Qy	230	ACSSKAVVSLRC--IACGVNLSRSOS-RIVGESALPGAWPMQVSLHVQNVHVCGSII	286
Db	535	QCLEDLSLILQCNYSKCGKLVQEVSPKIVGSGDSREGAWPVVALYFDDQVCGASLV	594
Qy	287	TPEWIVTAACHVEKPLNPNHWTAFAGILRQSFMYGAGYQVE-----KVISHPNYDSKT	341
Db	595	SRDLVSAACHCVYGRNMEPSKWKAVLGLHNASNL---TSPQIETRLIDQIVINPHYNKER	651
Qy	342	KNNDIALMKLQKPLTFNDLVKPVCLPNPMMLOPEOLCWI SGWGA TE EKGTSEVLNAAK	401
Db	652	KNNDIAMVHLEMKVNYTDYIQICLPEENQVFPFGRICSIAGWGALIYQGSTADVLQEAD	711
Qy	402	VLLIETORCNSRYVDNLIITPAMLCAGFLOGNVDSQGS GGP LVT SKNIWMLIGDTSW	461
Db	712	VPLLSNEKCOQOMPEYN-ITENMV CAGYEAGGVDSQGS GGP LMCQENRWLLAGVTSP	770
Qy	462	GSQCAKAYRPGVYGNVMVFTDWI	484
Db	771	GYQCALPNRPGVYARVPRFTEWI	793

Search completed: June 1, 2004, 14:39:03
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 14:15:33 ; Search time 134 Seconds
(without alignments)
1158.470 Million cell updates/sec

Title: US-09-615-285B-2
Perfect score: 2717
Sequence: 1 MALNSGSPPAIGFYENHGY.....VGNVMVPTDIYRQMRADG 492

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTRMBL25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_plant.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2711	99.8	492	Q96T73	Q96T73 homo sapien
2	2121	78.1	490	Q7TN04	Q7TN04 mus musculus
3	2083	76.7	490	Q920K3	Q920K3 rattus norv
4	1009	37.1	767	Q9DGR2	Q9DGR2 xenopus lae
5	881	32.4	453	Q812A6	Q812A6 mus musculus
6	751	27.6	537	Q9BYE1	Q9BYE1 homo sapien
7	736.5	27.1	581	Q9BYE2	Q9BYE2 mus musculus
8	731	26.9	471	Q8CFE0	Q8CFE0 mus musculus
9	725.5	26.7	558	Q86YM4	Q86YM4 homo sapien
10	676.5	24.9	405	Q96S86	Q96S86 homo sapien
11	675	24.8	445	Q8CU17	Q8CU17 mus musculus
12	664.5	24.5	455	Q8CDR0	Q8CDR0 mus musculus
13	662.5	24.4	371	Q8CU16	Q8CU16 rattus norv
14	649	23.9	326	Q7Z280	Q7Z280 brachydactyl
15	646.5	23.8	777	Q8CAN9	Q8CAN9 mus musculus
16	600.5	22.1	1111	Q80YN4	Q80YN4 rattus norv

17	567.5	20.9	624	6	Q95ME7	Q95ME7 oryctolagus
18	564.5	20.8	855	11	Q9DU17	Q9DU17 rattus norv
19	563.5	20.7	643	6	Q97506	Q97506 sus scrofa
20	560.5	20.6	624	11	Q9DAT3	Q9DAT3 mus musculus
21	558.5	20.6	422	4	Q8WVC1	Q8WVC1 homo sapien
22	553	20.4	624	11	Q91Y47	Q91Y47 mus musculus
23	544	20.0	1379	5	Q9VAN6	Q9VAN6 drosophila
24	543	20.0	845	13	Q9DGR1	Q9DGR1 xenopus lae
25	542	19.9	638	11	Q8ROP5	Q8ROP5 mus musculus
26	541	19.9	572	11	Q8BIK6	Q8BIK6 mus musculus
27	529	19.5	310	11	Q9QYZ9	Q9QYZ9 mus musculus
28	523	19.2	310	11	Q91XC4	Q91XC4 mus musculus
29	522.5	19.2	812	11	Q9ROW3	Q9ROW3 mus musculus
30	519.5	19.1	439	11	Q8BHM9	Q8BHM9 mus musculus
31	516	19.0	417	11	Q8VHJ4	Q8VHJ4 rattus norv
32	514	18.9	417	11	Q8VDV1	Q8VDV1 mus musculus
33	514	18.9	417	11	Q8VHK8	Q8VHK8 mus musculus
34	509.5	18.8	329	6	Q9GL10	Q9GL10 ovis aries
35	509.5	18.8	331	11	Q8RIA6	Q8RIA6 mus musculus
36	509.5	18.8	389	13	Q9PVX7	Q9PVX7 xenopus lae
37	508.5	18.7	279	11	Q7TNX3	Q7TNX3 mus musculus
38	508.5	18.7	417	11	Q8BZ10	Q8BZ10 mus musculus
39	506.5	18.6	331	11	Q8XK17	Q8XK17 mus musculus
40	505	18.6	328	11	Q8BUR6	Q8BUR6 mus musculus
41	504.5	18.6	257	11	Q8SZ04	Q8SZ04 mus musculus
42	504.5	18.6	320	13	Q7TDX2	Q7TDX2 xenopus lae
43	504.5	18.6	336	11	Q80YD8	Q80YD8 mus musculus
44	503.5	18.5	279	11	Q9QZ74	Q9QZ74 rattus norv
45	503.5	18.5	416	4	Q86T26	Q86T26 homo sapien

ALIGNMENTS

RESULT 1

Q96T73	PRELIMINARY;	PRT;	492 AA.
ID	Q96T73		
AC	Q96T73;		
DT	01-DEC-2001 (TRENBLrel. 19, Created)		
DT	01-DEC-2001 (TRENBLrel. 19, Last sequence update)		
DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)		
DE	Epitheliasin.		
GN	TPRS82.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21223025; PubMed=11322890;		
RA	Jacquinet E., Rao N.V., Rao G.V., Wang Z., Albertine K.H.,		
RA	Hoidal J.R.;		
RT	"Cloning and characterization of the cDNA and gene for human		
RT	epitheliasin."		
RL	Eur. J. Biochem. 268:2687-2699(2001).		
CC	-L- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.		
DR	EMBL; AF329454; AAKS3559.1; -		
DR	HSP; P00761; IANL.		
DR	GO; GO:0016020; C:membrane; IEA.		
DR	GO; GO:0004263; F:chymotrypsin activity; IEA.		
DR	GO; GO:0008233; F:peptidase activity; IEA.		
DR	GO; GO:0005044; F:scavenger receptor activity; IEA.		
DR	GO; GO:0004295; F:trypsin activity; IEA.		
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.		
DR	InterPro; IPR009003; Cys_Ser_trypsin.		
DR	InterPro; IPR002172; LBL_receptor_A.		
DR	InterPro; IPR001254; Peptidase S1.		
DR	InterPro; IPR001314; Peptidase S1A.		
DR	InterPro; IPR001190; Strc_receptor.		
DR	Pfam; PF00089; trypsin_1.		
DR	PRINTS; PR00722; CHYMOTRYPSIN.		
DR	SMART; SM00192; LDLA; 1.		
DR	SMART; SM00202; SR; 1.		

DR SMART; SM00020; Trp_Spc; 1.
DR PROSITE; PS01209; LDLRA 1; 1.
DR PROSITE; PS0068; LDLRA 2; 1.
DR PROSITE; PS0287; SRCR 2; 1.
DR PROSITE; PS0240; TRYPsin DOM; 1.
DR PROSITE; PS00134; TRYPsin_S1; 1.
DR PROSITE; PS00135; TRYPsin_S2; 1.
KW Hydrolyase, Protease; Serine protease.
SQ SEQUENCE 492 AA; 53863 MW; 38BA755BF276DADF CRC64;

Query Match 99.8%; Score 2711; DB 4; Length 492;
Best Local Similarity 99.6%; Pred. No. 1.1e-212;
Matches 490; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVEVHPAQYPSVPQYAPRVLTQA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVEVHPAQYPSVPQYAPRVLTQA 60

QY 61 SNPVVCTQKPSGFTVCTSKTKALCITLTGLTFLVGAALAAAGLLKFMGSKCSNSGIEC 120
Db 61 SNPVVCTQKPSGFTVCTSKTKALCITLTGLTFLVGAALAAAGLLKFMGSKCSNSGIEC 120

QY 121 DSSGTCINPSNWCDSVCHPCGDEENRCVRLYGPNFLOVYSSORKSMHWPVQDDWNNY 180
Db 121 DSSGTCINPSNWCDSVCHPCGDEENRCVRLYGPNFLOVYSSORKSMHWPVQDDWNNY 180

QY 181 GRAACRDMGYKNNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACKDNGYKNNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240

QY 241 CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVNQVHVCSSIIITPEWITVAACHVEK 300
Db 241 CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVNQVHVCSSIIITPEWITVAACHVEK 300

QY 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVIHSPNYDSKTKNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVIHSPNYDSKTKNDIALMKLQKPLTFNDL 360

QY 361 VKPCLPNPQWMLQPEQLCWISGSGATEBEKTKSEVLNAAKVLIEIETQCNRYVDNLI 420
Db 361 VKPCLPNPQWMLQPEQLCWISGSGATEBEKTKSEVLNAAKVLIEIETQCNRYVDNLI 420

QY 421 TPAMICAGFLOGNVDSQCGSGPLVTSKNIIWLLIGTWSGSGCAKAYRPGVYGNVWF 480
Db 421 TPAMICAGFLOGNVDSQCGSGPLVTSKNIIWLLIGTWSGSGCAKAYRPGVYGNVWF 480

QY 481 TDWIYRQMRADG 492
Db 481 TDWIYRQMRADG 492

RESULT 2
Q7TN04 PRELIMINARY; PRT; 490 AA.
AC Q7TN04;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Trpss2 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny K.D., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,
RA Fahey J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."; U.S.A. 99:16899-16903 (2002).
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC054348; AAHS4348.1; -- 546508028417665A CRC64;
SQ SEQUENCE 490 AA; 53525 MW; 546508028417665A CRC64;

Query Match 78.1%; Score 2121; DB 11; Length 490;
Best Local Similarity 78.4%; Pred. No. 3.7e-180;
Matches 385; Conservative 43; Mismatches 61; Indels 2; Gaps 2;

QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVEVHPAQYPSVPQYAPRVLTQA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVEVHPAQYPSVPQYAPRVLTQA 60

QY 61 SNPVVCTQKPSGFTVCTSKTKALCITLTGLTFLVGAALAAAGLLKFMGSKCSNSGIEC 120
Db 61 STSVITHTPKS -SGALCTSKSKSLCLALGTLTGAAGAAVLLWRFWDSNCSTSEMEC 119

QY 121 DSSGTCINPSNWCDSVCHPCGDEENRCVRLYGPNFLOVYSSORKSMHWPVQDDWNNY 180
Db 120 GSSGTCISSLLWCDGVAHCPNGEDENRCVRLYGPNFLOVYSSORKSMHWPVQDDWNNY 179

QY 181 GRAACRDMGYKNNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 180 GRAACKDNGYKNNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 239

QY 241 CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVNQVHVCSSIIITPEWITVAACHVEK 300
Db 240 CIECGVR -SVTKQSRIVGGLNASPGDPWQVSLHVGHVCGSGIITPEWITVAACHVEE 298

QY 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVIHSPNYDSKTKNDIALMKLQKPLTFNDL 360
Db 299 FUSSPRYWTFAGILRQSLMFYGSRRHQVEKVIHSPNYDSKTKNDIALMKLQKPLTFNDL 358

QY 361 VKPCLPNPQWMLQPEQLCWISGSGATEBEKTKSEVLNAAKVLIEIETQCNRYVDNLI 420
Db 359 VKPCLPNPQWMLQPEQLCWISGSGATEBEKTKSEVLNAAKVLIEIETQCNRYVDNLI 418

QY 421 TPAMICAGFLOGNVDSQCGSGPLVTSKNIIWLLIGTWSGSGCAKAYRPGVYGNVWF 480
Db 419 TPAMICAGFLOGNVDSQCGSGPLVTLKNGIWWLLIGTWSGSGCAKAYRPGVYGNVWF 478

QY 481 TDWIYRQMRAD 491
Db 479 TDWIYRQMRAN 489

RESULT 3
Q920K3 PRELIMINARY; PRT; 490 AA.
ID Q920K3
AC Q920K3;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Trpss2.
GN Trpss2.
OS Rattus norvegicus (Rat).

479 TDWIYQOMRAN 489

DB

RESULT 4

Q9DGR2

ID Q9DGR2 PRELIMINARY; PRT; 767 AA.

AC Q9DGR2;

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE XESP-2.

DE XESP-2.

GN

OS Xenopus laevis (African clawed frog).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

OC Xenopodinae; Xenopus.

OC NCBI_TaxID=8355;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=20363741; PubMed=10903452;

RX Yamada K., Takabatake T., Takeshima K.;

RT "Isolation and characterization of three novel serine protease genes from *Xenopus laevis*."

RT from *Xenopus laevis*."

RL Gene 252:209-216(2000).

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

CC EMBL; AB038497; BAB08217.1; -.

DR HSSP; P00766; 1CHG.

DR MEROPS; S01.049; -.

DR GO; GO:0004263; F.chymotrypsin activity; IEA.

DR GO; GO:0008233; F.peptidase activity; IEA.

DR GO; GO:0004295; F.trypsin activity; IEA.

DR GO; GO:0006508; P.proteolysis and peptidolysis; IEA.

DR InterPro; IPR009003; Cys_Ser_trypsin.

DR InterPro; IPR002172; LDL_receptor_A.

DR InterPro; IPR001254; Peptidase_S1.

DR InterPro; IPR001314; Peptidase_S1A.

DR Pfam; PF00057; ldl_recept_a; 4.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR PRINTS; PR00261; LDLRECEPTOR.

DR SMART; SM00192; LDLa; 8.

DR SMART; SM00020; Tryp_SPC; 1.

DR PROSITE; PS01209; LDLRA_1; 8.

DR PROSITE; PS00658; LDLRA_2; 2.

DR PROSITE; PS02440; TRYPSIN_DOM; 1.

DR PROSITE; PS00134; TRYPSIN_HIS; 1.

DR PROSITE; PS00135; TRYPSIN_SER; 1.

KW Hydroxylase; Protease; Serine protease.

QY SEQUENCE 767 AA; 86001 MW; E0566A3879DE36E CRC64;

Query Match 37.1%; Score 1009; DB 13; Length 767;

Best Local Similarity 50.9%; Pred. No. 7.2e-81;

Matches 189; Conservative 49; Mismatches 131; Indels 2; Gaps 2;

QY 120 CDSSGTCINPNCWGDVSHCPGGEENRCVRILYGNPFIQLQVYSSQRKSWHPVQDDWNEN 179

DB 395 CGSSVSVCLSSQWCDGVSDCPYGEDMSVSLYPADFLQVYSTSVSAMLFPVCSGYWDD 454

QY 180 YGRAACRDNGYKNFVSS-QGIVDDSGSTSPMKLINTSAGNVDIYKLYHSDACSKAVVS 238

DB 455 FGFSACQDFGYNGSSYNNRYDTLMSYPAGNYFKLYSGYWRSEKPYTSVQSSYCYSGNVVS 514

QY 239 LRCIACGVNLNSRQSRIRVGGESALPGAPWQVSLHVQNVHVCGGSIITPEWITVAACHV 298

DB 515 LHCISGVSNLSVSRIVGGTFANTGNWPQVNLQYITGVLCCGSIISPKWIVTAAACHV 573

QY 299 EKPLNPNPHTAPAGILRQSFMYGAGYQVEKVISHPNTDTSKTNNDIALMKLQKPLTFN 358

DB 574 YGYSYSSASGWRVPAGTLTKPSYNNASAYFERIIVHPGYKSYTYNDIALMKLRDEITFG 633

QY 359 DLVKPYCLNPNGMLOPEQLCWISGWCATEEKKTSEVLNAAKVLLETQRCNSRYVDN 418

DB 634 YTPQVCLPNSGMFWAGTTTWSGWSITVEGGSVSTYLOAYIPLIDSNVCQSYNYG 693

```
QY 419 LITPAMICAGFLQGNVDSGCGSGPLVTSKNNIWWLIGDTSWGSCKAYRPGVYGNM 478
Db 694 QITSSMICAGYLSGGVDTCGDSGGPLVYKRNKNTWHLVGDTSWGDCCARANKFVYGNVT 753
QY 479 VFTDWIYRQMR 489
Db 754 TFLWYISQMR 764

RESULT 5
Q812A6 PRELIMINARY; PRT; 453 AA.
AC Q812A6;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Transmembrane proteinase tmprss3.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RA Rao N.V., Rao G.N., Hoidal J.R.;
RT "Genomic Organization of Murine Transmembrane Proteinases.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF479687; AA033581.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR001190; S1cr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS0068; LDLRA_2; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS0134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
SQ SEQUENCE 453 AA; 49505 MW; 1BE7ECD6CB3DD894 CRC64;

Query Match 32.4%; Score 891; DB 11; Length 453;
Best Local Similarity 43.1%; Pred. No. 8.8e-70;
Matches 188; Conservative 63; Mismatches 159; Indels 26; Gaps 12;

QY 69 PKSPSGTCTSKTKKALCI-----TLTGLTFLVGAALAGLLWKPMSKNSGIECDSSG 124
Db 28 PVAPDGAQAQILSLPLKFFPIVIGIILALALAIGLIHF---DCSGK-YRCHSSF 83

QY 125 TCINPSNWCDSVSHCPGSGEDENRCVRLGPNFILQVYSQKSWHPVCQDDNENYGRAA 184
Db 84 KCIELTARCDGSDCKNAEDYRCVRVSGQRAALQVFTA--AAWTWCSDDKSHYAKIA 141

QY 185 CRDMGYKNNFYSSQGI-----VDDSGSTSMKLNLTAGNVDIYKLYHS-----DACSKAV 236
Db 142 CAQLGFP-SYVSSDLRLVDALDEEQDGFVSINLLSD-DKVTALHSHVYMRREGCTSGHV 199

QY 237 VSLRCIAGVNLNSRQRVIGGESALPGANPWVSLHVQNVHVCGSIIIPENVITAAH 296
Db 200 VTLKCSACGTRTGS--PRIVGNNSSLTQMPWQVSLQFQGYHLGCGSIIIPENVITAAH 257

QY 297 CVEKPLNPNFWHTAPAGILRQSFMPYGA-GYQVEKVI SHPNYDSKTKNNNDIALMKLQKPL 355
```

```
Db 258 CV-YDLXPKPSWTQVGLV--SLMDSVPVSHLVEKIIYHSKYKPKRLGNDIALMKLSEPL 314
QY 356 TFDNLVFPVCLPNPQMLQPEQLCWSGWGATEEKGKTSVYLNAAKVLIIETQRCNRYV 415
Db 315 TFDETIQICLPNSEENFPDGKLCWTSWGATEDGGDASPVLNHAAVPLISNKICNHRDV 374
QY 416 YDNLITPAMICAGFLQGNVDSGCGSGPLVTSKNNIWWLIGDTSWGSCKAYRPGVY 475
Db 375 YGGIISPSMLCAGYLKGVDSGCGSGPLVTSKNNIWWLIGDTSWGSCKAYRPGVY 434
QY 476 NMVFTDWIYRQMRAD 491
Db 435 RITSFLDWIHEQLERD 450

RESULT 6
Q89BYE1 PRELIMINARY; PRT; 537 AA.
AC Q89BYE1;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Mosaic serine protease.
GN MSPS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=21167393; PubMed=11267681;
RA Kim D.R., Sharmin S., Inoue M., Kido H.;
RT "Cloning and expression of novel mosaic serine proteases with and
without a transmembrane domain from human lung.";
RL Biochim. Biophys. Acta 1518:204-209 (2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AB048797; BAB39742.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.087; -.
DR GO; GO:0016021; C:integral to membrane; NAS.
DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR001190; S1cr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Protease; Serine protease.
SQ SEQUENCE 537 AA; 58102 MW; A39FF4E8816DAECF CRC64;

Query Match 27.6%; Score 751; DB 4; Length 537;
Best Local Similarity 32.0%; Pred. No. 4.2e-58;
Matches 173; Conservative 81; Mismatches 198; Indels 88; Gaps 14;

QY 5 SSSPPAIGYYENHGVQENPYPAQTVVPT-VYEVHPAQYVPSVP----- 50
Db 27 AGTPPORA-----SPAQAQSPAQSPAGTTPGRASPAQSPAGTTPGRASPGRASPAQ 78

QY 51 ----QYAPRLVTOA-SNPVVCVTPQKPSQG-----TVCTSKTKKALCIITLTGFLVG 97
Db 79 ASPAQAQSPAQSPARASPALISRSSSGRSSASASVTTSTPTVYLVRATPVGAVPTR 138
```


QY 98 AALA-----AGLLWKFMSKCSNSGIECDSSGTCINPNSMCDGVS 137
 Db 139 SSPARSAPATATRESVPQWQHTGIRYKEQRESCPKHVR-----CDGVV 185
 QY 138 HCPGGBENRCVLYGNFILOVYSSORKSWHPVQODNNENYGRAACRDMGYKXNFYSS 197
 Db 186 DCKLSDELGCVRFDWPKSLKIYSGSSHQWLPICSSNNWDSYSEKTCRQLQFESAHRRT 245
 QY 198 Q-GTVDDSGSTSMKLNSTAGNDVYKLYHSDACSKAVVSRUCIACGVNLNSRSQRI 256
 Db 246 EVAHRDPANFESILRYNST-----IQESLHRSCHPCSORVISLQCSHGCLR---AMTGR 296
 QY 257 VGGESALUPGAPWQVSLHVQNVHVCSSGIITPENVITAAHCV-----EKPLNPNHWTAFAGILRQSEM 312
 Db 297 VGGALADSQWFWQVSLHFGTHICGGTLIDAQWVLTAAHCFVTRKVLG---WKVYA 353
 QY 313 GILRQSMFYCAGVQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDLVKPCVLEPNGMV 372
 Db 354 GTSNLHQLPENA--SIABIIINSNYTDEDDYDIALMRLSKPLTSLSAHHPACLPHGQT 411
 QY 373 LQPEQLCNIWSGATEE-KGKTSEVLNAKVLLETQRCNSRYVYDNLITPAMICAGFLQ 431
 Db 412 FSLNETCWTGFGKTRTDDKTSPLREVQVNLIDFKKNDYLVDSYLTTPRMWCAGLD 471
 QY 432 GNVDSCQDGGPLVTSKNINWILGDTSMGSCAKAYRFGVGNVNVFTDWIYRQ 491
 Db 472 GGRDSCQDGGPLVCEQNNRWLAGVTSWGTGCGQRNKGVTYKTVTEVLPWIYSKM 531

RESULT 7
 Q9BYE2 PRELIMINARY; PRT; 581 AA.
 AC Q9BYE2
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Membrane-type mosaic serine protease.
 GN MSPL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=21167393; PubMed=11267681;
 RA Kim D.R., Sharmin S., Inoue M., Kido H.;
 RT "Cloning and expression of novel mosaic serine proteases with and
 without a transmembrane domain from human lung.";
 RL Biochim. Biophys. Acta 1518:204-209 (2001).
 DR EMBL; AB048796; BAB39741.1; -;
 DR HSP; P00763; IDPO.
 DR MEROPS; S01.087; -;
 DR GO; GO:0016021; C:integral to membrane; NAS.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
 DR InterPro; IPR009003; Cys_ser_trypsin.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR Pfam; PF00057; ldl_recept_a; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDla; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00202; Tryp_Spc; 1.
 DR PROSITE; PS00287; SRCR 2; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 581 AA; 62689 MW; 4DABE24D7D5BA44 CRC64;

Query Match 27.1%; Score 736.5; DB 4; Length 581;
 Best local similarity 22.3%; Pred. No. 9.1e-57;
 Matches 171; Conservative 78; Mismatches 198; Indels 82; Gaps 14;
 QY 5 SGSPALGPYENHGYQENPYPAQPTVTVVTVVHVAQYYP-----46
 Db 62 AGTPPGRA-----SPGRASPAQAS-----PARASPALASLSRSSSSGRSSARS 104
 QY 47 -----SPVPOYAPRVLTAQSNPVVCTQPKSPSGTCTCTKKALCITLT-----L 91
 Db 105 ASVTTSPTVYLVRAVTVGAVPIRSSPARSAPATATRESPTSLPKFTWREGQKQLPLI 164
 QY 92 GTFVLGAALAAAGLL-----WKFMGSKCSNSGI-ECDSSGTCINPNSMCDGVSCHPCGEDE 145
 Db 165 GCVLLALIALVSLIILFQFWQ-----GHTGIRHKEQRESCPKHVRCDGVVDCCLKSDE 218
 QY 146 NRCVRLVGPFILOVYSSORKSWHPVQODNNENYGRAACRDMGYKXNFYSSQ-GIVDSD 204
 Db 219 LGCVRFWDKSLKTIYSGSSHQWLPICSSNNWDSYSEKTCRQLQFESAHRTEVAHRDFA 278
 QY 205 GSTSMFKLNSTAGNDVYKLYHSDACSKAVVSRUCIACGVNLNSRSQRIYVGGESALP 264
 Db 279 NSFSILRYNST-----IQESLHRSCHPCSORVISLQCSHGCLR---AMTGRIVGGALAD 329
 QY 265 GAMPWQVSLHVQNVHVCSSGIITPENVITAAHCV-----EKPLNPNHWTAFAGILRQSEM 320
 Db 330 SKMPWQVSLHFGTHICGGTLIDAQWVLTAAHCFVTRKVLG---WKVYAGTSMHOL 386
 QY 321 FYGAGVQVEKVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPCVLEPNGMVLPQELCW 380
 Db 387 PEAA--SIABIIINSNYTDEDDYDIALMRLSKPLTSLSAHHPACLPHMGQTFSLNETCW 444
 QY 381 ISGWGATEE-KGKTSEVLNAKVLLETQRCNSRYVYDNLITPAMICAGFLQNVDSGCG 439
 Db 445 ITGFGKTRTDDKTSPLREVQVNLIDFKKNDYLVDSYLTTPRMWCAGLDGGRRDSCG 504
 QY 440 DSGGGLVTSKNINWILGDTSMGSCAKAYRFGVGNVNVFTDWIYRQ 498
 Db 505 DSGGGLVCEQNNRWLAGVTSWGTGCGQRNKGVTYKTVTEVLPWIYSKM 553
 RESULT 8
 Q8CFE0 PRELIMINARY; PRT; 471 AA.
 AC Q8CFE0
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Similar to mosaic serine protease (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
 RA Strauberg R.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC042878; AAH42878.1; -;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0005044; F:scavenger receptor activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_ser_trypsin.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srcr_receptor.
 DR Pfam; PF00057; ldl_recept_a; 1.
 DR Pfam; PF00530; SRCR; 1.
 DR Pfam; PF00089; trypsin; 1.

DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART; SM00202; SR; 1
DR SMART; SM00202; TRYP_SPC; 1.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Protease.
FT NON IER
SQ SEQUENCE 471 AA; 52535 MW; ED58CFB6B7C3CC4 CRC64;

Query Match 26.9%; Score 731; DB 11; Length 471;
Best Local Similarity 35.8%; Pred. No. 2.1e-56;
Matches 166; Conservative 75; Mismatches 183; Indels 44; Gaps 13;

QY 47 SPVPEQYAPRVLTQASNPVCTQKPSGTVCT-----SKTKAL-----CITL 89
DB 19 SPTRVVLVRATPGVAVPIRASPAPATRAFPSPGLSPFKFSQETQRLPLIGCVIL 78
QY 90 TLGTFVLVGAALAAAGLLWKMGKCSNSGIEC-DSSGTCINPSNWCDDGVSHCPGGEDNRC 148
DB 79 -----LISLVISILLIFYFWR---GHTGIKYEPLSCFPIHVRCDGVVDCMKSDSLGC 130
QY 149 VRLYGPNFIQVYSSQKSWHPVCCDDNENYGRAACRDMGYKNNFYSSQGIIVDDSGTS 208
DB 131 VRFDWKDKSLKYVSGSGEWLPVCSNNDWTDKSKTCCQLGF-DSAYETTEVAHRDITSS 189
QY 209 FMKLNTSAGNVDIYKLYHSDACSKAVVSLRACIACGVNLNNSRQSRIVGGESALPGAMP 268
DB 190 FL---LSEYNTTIQESLYRSQ-CPERRVYSLQSCRCGLR---AMTGRIVGGALTSSEKWP 242
QY 269 MOWSLHVQNVHVCGGSIITPEWIVTAACHV-----EKPLNPNWHTAPAGILRQSFYGA 324
DB 243 MOWSLHFGTHICGGTILDAQWVLTAAHCFVTRKLEGG---WKYVAGTSLNHLQLEAA 299
QY 325 GYQVEKVIHSHPNYDSKTNNDIALMKLQKPLTFNDLVKPVCLPMPGMLQPEQLCWISGW 384
DB 300 --SISQIIINGNTYDEQDDYDIALRLSKPLTSLSAHHPALCPMHEGQTFLGNETCWTGF 357
QY 385 GATEERK-KTSEVLNAAKVLIIETQRCNSRVYVDNLITPAMICAGFLQGNVDSQCGSGG 443
DB 358 GKTKEDEKTSFPLREVQVNLIDFKKNDYLVDSYLTFRMWCAGDLRGGEDSCQGSQGG 417
QY 444 PLVTSKNNIWLIGTWSGSGCAKAYRPGVYGNVMVFTDNIYRQMRAD 491
DB 418 PLVCEQNNRWYLAGVTSWGTGCGQKPKGVYTKVTEVLPFIYRRKMESE 465

RESULT 9
Q86YM4 PRELIMINARY; PRT; 558 AA.
AC Q86YM4; (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Transmembrane protease serine 6.
GN TMPSR56.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RA Park T.J.; Park W.J.;
RP "Homo sapiens transmembrane protease, serine 6 (TMPSR56) mRNA."
RT Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AV190317; AAC38062.1;
DR GO; GO:0016020; Cmembrane; IEA.
DR GO; GO:0004263; Fchymotrypsin activity; IEA.
DR GO; GO:0008233; Fpeptidase activity; IEA.
DR GO; GO:0005044; Fscavenger receptor activity; IEA.
DR GO; GO:0004295; Ftrypsin activity; IEA.
DR GO; GO:0006508; Pproteolysis and peptidolysis; IEA.

DR InterPro; IPR009003; Cys_Ser_tryptase.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Strc_receptor.
DR Pfam; PF00089; tryptase; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00202; SR; 1.
DR SMART; SM00202; TRYP_SPC; 1.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Protease.
SQ SEQUENCE 558 AA; 60432 MW; 4AC817FCD70D7017 CRC64;

Query Match 26.7%; Score 725.5; DB 4; Length 558;
Best Local Similarity 32.2%; Pred. No. 8.2e-56;
Matches 171; Conservative 81; Mismatches 197; Indels 82; Gaps 15;

QY 5 SGSPALGAYENHGQYQENPYPAQPTVVPVYVNHQAQYYP-----L 46
DB 62 AGTTPGPA-----SPGRAPPAQAS-----PARASPALASLSRSSSSGRSSARS 104
QY 47 -----SPVPEQYAPRVLTQASNPVCTQKPSGTVCTSKTKALCITLT-----L 91
DB 105 ASVTTSPTRVVLVRATPGVAVPIRSPAPATRAFPSPGLSPFKFWREGOKQLPLI 164
QY 92 GTFLVGAALAAAGLL-----WKFMGSKCSNSGIEC-DSSGTCINPSNWCDDGVSHCPGGED 145
DB 165 GCVLIIILVLSLILFQWQ-----GYTGIRYKEQRESCEPHEAVERDGVDCKLKSD 218
QY 146 NCVSLYGNFNTLOVYSSQKSWHPVCCDDNENYGRAACRDMGYKNNFYSSQ-GIIVDDS 204
DB 219 LCCVDFDWKDKSLKYVSGSGEWLPVCSNNDWTDKSKTCCQLGF-DSAYETTEVAHRDFA 278
QY 205 GSTSPMKLNTSAGNVDIYKLYHSDACSKAVVSLRACIACGVNLNNSRQSRIVGGESALP 264
DB 279 NFSFSLRYNST-----IQESLHRS-CPERSYVSLQSCRCGLR---AMTGRIVGGALASD 329
QY 265 GAWPQVSLHVQNVHVCGGSIITPEWIVTAACHV-----EKPLNPNWHTAPAGILRQSF 320
DB 330 SKWPMQVSLHFGTHICGGTILDAQWVLTAAHCFVTRKLEGG---WKYVAGTSLNHLQ 386
QY 321 FYGAGYQVEKVIHSHPNYDSKTNNDIALMKLQKPLTFNDLVKPVCLPMPGMLQPEQLCW 380
DB 387 FEAA--SIAETIINSNYTDEDDYDIALMRLSKPLTSLSAHHPALCPMHEGQTFLSNETCW 444
QY 381 ISGWGATEE-KGKTSEVLNAAKVLIIETQRCNSRVYVDNLITPAMICAGFLQGNVDSQCG 439
DB 445 ITGFGKTRTDDKTSPFLREVQVNLIDFKKNDYLVDSYLTFRMWCAGDLRGGEDSCQGS 504
QY 440 DSGGLVTSKNNIWLIGTWSGSGCAKAYRPGVYGNVMVFTDNIYRQMR 490
DB 505 DSGGLVCEQNNRWYLAGVTSWGTGCGQKPKGVYTKVTEVLPFIYRRKMESE 555

RESULT 10
Q95E86 PRELIMINARY; PRT; 405 AA.
AC Q95E86; (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 25, Last annotation update)
DE Similar to transmembrane protease, serine 4 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RA SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.


```

DE Transmembrane protease.
GN TMPSRS
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK029714; BAC26577.1; -.
DR MGD; MGI:1933407; Tmpres5.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Ssr_receptor.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0287; SSCR_2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
SQ SEQUENCE 455 AA; 49669 NW; B222BE2E7503C74B CRC64;

Query Match 24.5%; Score 664.5; DB 11; Length 455;
Best Local Similarity 35.0%; Pred. No. 1.7e-50;
Matches 145; Conservative 66; Mismatches 176; Indels 27; Gaps 9;

QY 86 CITLTLGTLVGAALAAGLMKFGSKCSNGIECDSGTCINFNWCDGVSHCPGGEDE 145
DB 49 CVILGVGLGAGIASLLVLYLWPPASPS-----ISGTLQEB-----EMTLNCPGVSR 99

QY 146 NRCV-----RLYGNFLOVYSSQKSWHPVQDDNENYGRAACRDMGY-KNNFY 195
DB 100 BELLPSLPTKTVSRINGEDLLLOVQVRAAPDMLLVCHEGNWPALGMHCKSLGHIRLTQH 159

QY 196 SSQGIWDD--SGSTSFMKLNT-SAGNVDIYKLYHSDACSKAVVSLRCIACGVNLSNRQ 253
DB 160 KAVNLSDIKLNRSQEFALSRPGL-VEESWKPSANCPGRIVSLKSCCGAR---PLA 215

QY 254 SRVGGESALPGAWPKQVSLHVNHYVCGGSIITPEWIVTAAHCVK-PLNNPWHWTAF 312
DB 216 SRVGGQAVASGRWPQASVNLGSRHTCGSVLAPYVWVTAHCHMYFRLSLSSWRVHA 275

QY 313 GILRSFMFYGAGYQVEKVISHPNYDSKTKNNDIAMKLOKPLTFNDLVKPVCLPNPQM 372
DB 276 GLVSHGAVRHQGTWVEKIIPHPLYSAQNHVDYVALLQLRTPINFSDTVGAVCLPAKEQH 335

QY 373 LOPEQLCWISGWGATE-EKGKTSVNLNAKVLLIETQRCNSRYVDNLITPAMICAGFLQ 431
DB 336 FFWGSCQVSGWGHTDPSHTSSDTLQDTWVPLSTDLNCSNCSMYSGALTHRMCLAGYLD 395

QY 432 GNVDSQCGSGGLPVTSKNNIWMKLIQDTSWGSCKAKAYRPGVYGNVMVFTDWIY 485
DB 396 GRADACQGSGLPVCPGSDTHLVGVVSWGRGCAEPNRPVYAKVAEFLDWIH 449

RESULT 13
Q8CJ16 PRELIMINARY; PRT; 371 AA.
AC Q8CJ16;
DE 01-MAR-2003 (TrEMBLrel. 23, Created)

Query Match 24.4%; Score 662.5; DB 11; Length 371;
Best Local Similarity 37.7%; Pred. No. 1.9e-50;
Matches 133; Conservative 57; Mismatches 148; Indels 15; Gaps 6;

QY 141 GGDENRCVRLYGNPFILOVYSSQKSWHPVQDDNENYGRAACRDMGYKNNFYSSQ-- 198
DB 20 GQQQPISRFRINGEDLLLEQVRAAPDMLLVCHEGNWPALGMHICQSLGY---FRLTQHK 76

QY 199 ----GIVDSGSGTSMKLNLSAGNVDIYKLYHSDACSKAVVSLRCIACGVNLSNRQ 254
DB 77 AVNLSDIKLNRSQEFALSRPGL-VEEAWQFSTNCPGRIVSLKSCCGAR---PLAS 132

QY 255 RIVGGESALPGAWPKQVSLHVNHYVCGGSIITPEWIVTAAHCVK-PLNNPWHWTAF 313
DB 133 RIVGGQAVASGRWPQASVNLGSRHTCGSVLAPYVWVTAHCHMYFRLSLSSWRVHA 192

QY 314 ILRQSFMYGAGYQVEKVISHPNYDSKTKNNDIAMKLOKPLTFNDLVKPVCLPNPQM 373
DB 193 LVSHSAVRHQGTWVEKIIPHPLYSAQNHVDYVALLQLRTPINFSDTVSAVCLPAKEQH 252

QY 374 QPQLCWISGWGATE-EKGKTSVNLNAKVLLIETQRCNSRYVDNLITPAMICAGFLQ 432
DB 253 PQSGQCVSGWGHTDPSHTSSDTLQDTWVPLSTDLNCSNCSMYSGALTHRMCLAGYLD 312

QY 433 NVDSQCGSGGLPVTSKNNIWMKLIQDTSWGSCKAKAYRPGVYGNVMVFTDWIY 485
DB 313 RADACQGSGLPVCPGSDTHLVGVVSWGRGCAEPNRPVYAKVAEFLDWIH 365

RESULT 14
Q7Z280 PRELIMINARY; PRT; 326 AA.
AC Q7Z280;
DE 01-JUN-2003 (TrEMBLrel. 24, Created)
DE 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SI:dz69G10.3 (Novel protein similar to human transmembrane protease,
DE serine 3 (TMPRSS3)) (Fragment).

```

GN SI:D269G10.3.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Corby N.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL672083; CAD61105.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00668; LDLRA_2; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPsin DOM; 1.
DR PROSITE; PS00134; TRYPsin HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR Transmembrane; Protease.
KW Transmembrane; Protease.
FT NON_TER 1
SQ SEQUENCE 326 AA; 35561 MW; 8D6F12214393CDB2 CRC64;

Query Match 23.9%; Score 649; DB 13; Length 326;
Best Local Similarity 36.9%; Pred. No. 2.6e-49;
Matches 140; Conservative 50; Mismatches 121; Indels 68; Gaps 8;

QY 110 GSKCSNGIECDSSGTCINPSNWCDCGVSHCPGEGEDNRCVRLYGNFILQVYSSORKSWH 169
DB 1 GLRSCGKFCVSVSRVCSIRNAVCDGVQDCRDGEDLNCVRVSGSHSLQVFG--RGLWR 58

QY 170 PWCDDWNYGPAACRDGKXNFYSSQGIYDDSGTSFMTKNTSAGNVDIYKXLYHSD 229
DB 59 TVCSGWDQSLSLACRQLG----- 79

QY 230 ACSKAVSLRCLACGNLNSRSRQSRIVGSGALPGAWPQVSLHVQNVHVCSSIIITPE 289
DB 80 ---SRSAIS-----SRIVGNVSKSQVFPWQVSLHYQNYLQCGSIISHS 121

QY 290 WIVTAAHCVEKPLNPNHWTAPAGILRQSFMYGAGYQVEKVIISHPNYDSKTKNNDIALM 349
DB 122 WILTAACHV--FGPAQVPLMDVYAGLNLPL--LSKAEHRSVEKIIYHANFRSKSFSDIALI 179

QY 350 KLOKPLTFNDLVKPLNPGMMLOPEQLCWISGWCATBEKGTSEVLNAAKVLLIETOR 409
DB 180 KLLPLTFNDQIAPICLPNKGESFKNQMCCLISGWCATVDSGETSLSLHVAQVPLLSNKE 239

QY 410 CNSRYVNDLIITPAMICAGFLQNVDSQCGDGGPLVTSKNNIWMILGTSWGSCAKAY 469
DB 240 CRKLG-----LTNNVCTEFLRG-VGTQCGDSGGPL-ACQGSANTLVGTGSDWDCGKYN 292

QY 470 RGVYGNVMTDVIYROM 488
DB 293 KPGIYTSISEALTWIOEQM 311

RESULT 15
Q8CAN9

ID Q8CAN9 PRELIMINARY; PRT; 777 AA.
AC Q8CAN9
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Protease (Fragment).
GN PRSS7 OR A130097D21RIK.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RC MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK038356; BAC29973.1; -.
DR MGD; MGI:1197523; Pss87
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR00859; CUB.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000998; MAM_domain.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00629; MAM; 1.
DR Pfam; PF00530; SRCR; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 1.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00137; MAM; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00668; LDLRA_2; 1.
DR PROSITE; PS00740; MAM_1; 1.
DR PROSITE; PS00650; MAM_2; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPsin DOM; 1.
DR PROSITE; PS00134; TRYPsin HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
FT NON_TER 1
SQ SEQUENCE 777 AA; 87314 MW; A18E2F4ECF06D3A8 CRC64;

Query Match 23.8%; Score 646.5; DB 11; Length 777;
Best Local Similarity 30.5%; Pred. No. 1.4e-48;
Matches 164; Conservative 92; Mismatches 196; Indels 85; Gaps 20;

QY 18 HGYQENPYPAQPTWPT-----VVEV-HPAQYTPSPVPQYAPRVLTQAS---- 61
DB 253 NGICSQSPY-EPILVTFPPPELFTDCGGPFELWEPNSTSS--PNFPDKYPNQASCIWN 309

QY 62 -----NPV-----VCTQPKSPSGIVCTSKTKAL 85
DB 310 LNAQRGKNIQLHFQFDLENINDVVEVRDGGEDFSLLLAVYTG-GEVKDLFTSTNRMTV 368

QY 86 CITLTGTFVIG--AALAGLLWFMG--SKCSNGIECDSSGTCINPSNWCDCGVSHCPG 141
DB 369 IFTTNMTRRKGFKANFTSGY---YLGIPFCQDDEFQC-KDGNICPLGNLCLDSVPHCRD 424

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 1, 2004, 14:45:30 ; Search time 3778 Seconds
(without alignments)
3988.881 Million cell updates/sec

Title: US-09-615-285B-2
Perfect score: 2717
Sequence: 1 MALNSGPPAIGPYENHGY.....VYGNVMVFTDIYRQMRADG 492

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/sgn2_1/USPTO_spool/US09615285/tunat_01062004_132133_19962/app_query.fasta_1.647
-DB=EST -OPMT=fastap -SUFFIX=first -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09615285 @CGN 1 1 3437 @tunat_01062004_132133_19962 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_mam:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rtd:*
- 26: em_gss_phg:*
- 27: em_gss_vrl:*
- 28: gb_gesl:*

29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2674	98.4	1749	11	BC035623	Homo sapi
2	2291	84.3	1452	29	AY419481	Homo sapi
3	1770	65.1	1446	29	AY419483	Mus muscu
4	1770	65.1	1452	29	AY419482	Pan trogl
5	1237.5	45.5	866	12	BI761763	603046751
6	1198	44.1	781	13	BX508867	DKF3P686C
7	1100	40.5	603	13	BX479625	DKF3P686E
8	1095.5	40.3	621	12	BI414727	602991112
9	1088	40.0	604	14	CD724002	CJ28a11.Y
10	1084	39.9	678	14	CF147359	UI-HF-C80
11	1066	39.2	591	12	BM747154	K-EST0031
12	1046.5	38.5	736	12	BI454294	603171527
13	1040	38.3	767	14	CB989588	AGENCOURT
14	1028	37.8	644	14	CF147031	UI-HF-C80
15	1000	36.8	622	14	CF147378	UI-HF-C80
16	961.5	35.4	597	12	BI454495	603171459
17	914.5	33.7	796	12	BI556978	603238439
18	914.5	33.7	944	13	BQ228388	AGENCOURT
19	895	32.9	1034	13	BQ241270	603323236
20	894	32.9	910	12	BG242090	602354634
21	892.5	32.8	814	10	BF675496	602138474
22	890.5	32.8	648	10	BF102443	601751865
23	885	32.6	486	13	EX099916	EX099916
24	873	32.1	480	12	BM771081	K-EST0054
25	870.5	32.0	961	10	BF584569	602098278
26	858	31.6	892	13	BUS24316	AGENCOURT
27	857.5	31.6	1137	12	BI687501	603315055
28	844	31.1	737	10	BF168078	601776336
C	838	30.8	529	9	AW058537	wx23b08.x
29	833	30.7	592	13	BUS06558	603737747
30	830	30.5	816	13	BUS97449	603536894
31	830	30.3	639	12	BM784118	K-EST0062
32	823	30.3	725	12	BM784096	K-EST0062
33	823	30.3	528	10	BF076876	226635.MA
34	813	29.9	838	12	BI144021	602908124
35	813	29.9	465	10	BES57738	7947f08.x
C	806	29.7	465	10	BES57022	OV3-FT021
C	795	29.3	497	10	BES57022	OV3-FT021
36	794	29.2	744	14	CB600520	AGENCOURT
37	793	29.2	790	10	BF144549	601790193
38	790	29.1	703	10	BE376601	601226430
39	786	28.9	915	12	BI143513	602307696
40	786	28.9	783	12	BGI76274	602338030
41	783	28.8	846	12	BG870706	602791694
42	781.5	28.8	846	12	BG870706	602791694
43	781.5	28.8	846	12	BG870706	602791694
44	781	28.7	693	10	BES16896	601666437
C	764	28.1	450	9	AI393270	tg09d11.x

ALIGNMENTS

RESULT 1
BC035623
LOCUS
DEFINITION
Homo sapiens, similar to kallikrein B, plasma (Fletcher factor) 1,
clone IMAGE:5187060, mRNA.
1749 bp mRNA linear HTC 20-SEP-2002
ACCESSION
BC035623.1 GI:23242886
VERSION
BC035623
KEYWORDS
HTC.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1749)

AUTHORS	Strausberg, R.		QY	101	AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys	120
TITLE	Submitted (31-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			Db	417	GCGCTGGCCCTACTCTGGAGTTTCATGGGCGAGCAAGTGTCTCAACTCTGGATAGAGTGC
REMARK COMMENT	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgabs-remail.nih.gov Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc_mgc@hghri.nih.gov Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granitsis, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaipi, R., Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Teurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.		QY		121	AspSerSerGlyThrCysIleAenProSerAenTyrCysAspGlyValSerHisCysPro
				477	GACTCTCAGGTACCTGCATCAACCCCTCTAACTGGTGTGATGGCGTGTCACTGCCCC	536
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAX Plate: 68 Row: 1 Column: 6 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 14602458 This clone has the following problem: frame shifted.		QY	141	GlyGlyGluAspGluAenArgCysValArgLeuTyrGlyProAenPheIleLeuGlnVal	160
				537	GGCGGGAGGACGAGAAATCGGTGTGTTCGGCTCTACGGACCAAACTTCTATCTTCAGATG	596
source	1. 11749		QY	161	TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAenTyr	180
	/organism="Homo sapiens"			597	TACTCATCTCAGAGGAAGTCTGGCACCTGTGTGCCAAGACGACTGGAACGAGAACTAC	656
FEATURES	Location/Qualifiers		QY	181	GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAenPheTyrSerSerGlnGlyIle	200
	/mol_type="mRNA"			657	GGCGGGGGCGCTGCAGGACATGGCTATAAGAAATAATTTTACTCTAGCAAGGAATA	716
source	/db_xref="taxon:9606"		QY	201	ValAspAspSerGlySerThrSerPheMetLysLeuAenThrSerAlaGlyAenValAsp	220
	/clone="IMAGE:5187060"			717	GTGGATGACAGCGGATCCACGAGCTTTATGAACTGAACACAAAGTCCGCGCAATGTGAT	776
FEATURES	1. 11749		QY	221	IleTyr-LysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuAr	240
	/note="Vector: pCMV-SPORT6"			777	ATCTATAAATAAC-----CACAGTGTGCTGTCTTCAAAGCAGTGGTTCTTTTACG	830
source	/tissue types="Colon, kidney, stomach, adult, whole pooled"		QY	240	GlyGlyGlySerIleLeuThrProGluTyrPheValThrAlaAlaHisCysValGlyLys	260
	/clone="NIH MGC_11f"			831	CTGTATAGCTTGGCGGGTCAACTTGAATCAAGCCCGCAGAGCAGGATCGTGGCGGTGA	890
FEATURES	Alignment Scores:		QY	260	uSerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAenValHisVa	280
	Pred. No.: 4.19e-241 Length: 1749			891	GAGCGCGCTCCCGGGGGCTGGCCCTGGCAGGTGAGCTGCAGCTCCAGAACGTCACGT	950
source	1. 11749		QY	280	lCysGlyGlySerIleLeuThrProGluTyrPheValThrAlaAlaHisCysValGlyLys	300
	11. 11749			951	GTGGCGAGGCTCCATCATCACCCCGAGTGGATGCTGCAGCCGCCCTCCAGTGGGAAAA	1010
FEATURES	US-09-615-285B-2 (1-492) x BC035623 (1-1749)		QY	300	sProLeuAenAsnProTyrHisThrAlaPheAlaGlyIleLeuArgGlnSerPheMe	320
				1011	ACCTCTTAACAATCCATGGCATTTGACGGCATTTTGGCGGGATTTTGAGACAACTCTTTCAT	1070
source	1. 11749		QY	320	tPheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAenTyrAspSerLys	340
	11. 11749			1071	GTTCCTATGGAGCGCGATACCAAGTAGAAAAAGTAGATTTCTCATCAAATTTATGACTCCA	1130
FEATURES	US-09-615-285B-2 (1-492) x BC035623 (1-1749)		QY	340	sThrLysAenAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAspLe	360
				1131	GACCAAGAACATGACATTCGCTGATGAACTGCAGAACGCTCTGACTTTCAACGACCT	1190
source	1. 11749		QY	360	uValLysProValCysLeuProAenProGlyMetMetLeuGlnProGluGlnLeuCysTr	380
	11. 11749			1191	AGTGAACACAGTGTGTCTGCCCAACCCAGGCGATGATGCTGCAGCCAGAACAGCTCTGCTG	1250
FEATURES	US-09-615-285B-2 (1-492) x BC035623 (1-1749)		QY	380	pIleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAenAlaAl	400
				1251	GATTTCCGGGTGGGGGGCCCGCAGGAGAAAGGAAAGACCTCAGAAGTGTGGAACGCTGC	1310
source	1. 11749		QY	400	AllysValLeuLeuLeuGluThrGlnArgCysAenSerArgTyrValTyrAspAenLeuI	420
	11. 11749			1311	CAAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGTCTATGACAACTGAT	1370
FEATURES	US-09-615-285B-2 (1-492) x BC035623 (1-1749)		QY	420	eThrProAlaMetIleCysAlaGlyPheLeuGlnCysAenValAspSerCysGlnGlyAs	440
				1371	CACACGCGCATGATCTGTGCGGCTTCTGCGGGAGAACGTCGATTTCTTCCAGGGGTGA	1430
source	1. 11749		QY	440	pSerGlyGlyProLeuValThrSerLysAenAsnIleTyrTrpLeuLeuGlyAspThrSe	460
	11. 11749			1431	CAGTGGAGGGCTCTGGTCACTTCGAGAAACAATATCTGTGGTGTGCTGATAGGGGATACAAG	1490


```

QY 460 rTgGlySerGlyCysAlaValArgProGlyValTyrGlyAsnValMetValph 480
Db 1491 CTGGGGTCTGGCTGGCCAAAGCTTACAGACGAGGTACGGGAATGATGATGATT 1550
QY 480 eThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1551 CACGAGCTGGATTATCGACAAATGAGGGCAGACGGC 1587

RESULT 2
AY419481 1452 bp DNA linear GSS 17-DEC-2003
LOCUS Homo sapiens TMPRSS2 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY419481
VERSION AY419481.1 GI:39775438
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1452)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1452)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..1452
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1452
/gene="TMPRSS2"
/locus_tag="HCM6905"
gene
ORIGIN
Alignment Scores:
Pred. No.: 4,44e-205 Length: 1452
Score: 2291.00 Matches: 424
Percent Similarity: 87.60% Conservative: 0
Best Local Similarity: 87.60% Mismatches: 60
Query Match: 84.32% Indels: 0
DB: 29 Gaps: 0

US-09-615-285B-2 (1-492) x AY419481 (1-1452)
QY 6 GlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyrGlnProGluAsnPro 25
Db 1 GGGTACACACAGCTATTGGACCTTACTATGAAACCATGGATACCAACCGGAAACCCC 60
QY 26 TyrProAlaGlnProThrValValProThrValTyrGluValHisProAlaGlnTyrTyr 45
Db 61 TATCCGCGACAGCCCACTGTGTGCCACTGTCTACGAGGTGATCCGGCTCAGTACAC 120
QY 46 ProSerProValProGlnTyrAlaProArgValLeuThrGlnAlaSerAsnProValVal 65
Db 121 CCGTCCCCCGTCCCGACGTACGCCCGGAGGTCCTGACGCAGGCTCCCAACCCCGCTC 180
QY 66 CysThrGlnProLysSerProSerGlyThrValCysThrSerLysThrLysLysAlaLeu 85

```

```

181 TGCACGCAGCCCAAAATCCCCATCCGGGACAGTGTGCACCTCAAGACTAAGAAAGCACTG 240
QY 86 CysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeuAlaGlyLeuLeu 105
Db 241 TGCATACCTTGACCTGGGACCTTCTCGTGGAGCTCGCTGGCCGTGCCTACTC 300
QY 106 TrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCysAspSerSerGlyThr 125
Db 301 TGGAGTTTCATGGGACGAGTGTCTCAACTCTGGGATAGAGTGCAGTCTCTCAGGTACC 360
QY 126 CysIleAsnProSerAsnTrpCysAspGlyValSerHisCysProGlyGlyGluAspGlu 145
Db 361 TGCATCAACCCCTTAACCTGTGTGTCGCTGCACACTGCCCGCGGGAGGAGCAG 420
QY 146 AsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnValTyrSerSerGlnArg 165
Db 421 AATCGGTGTGTTCGCTCTACGGACCAAACTTCTCTTCAGGTGTACTCATCTCAGAGG 480
QY 166 LysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyrGlyArgAlaAlaCys 185
Db 481 AAGTCTCTGGCACCTGTGTGCCAAGACGACTGGAACGAGAACTACGGCGCGGGCGCTGC 540
QY 186 ArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIleValAspAspSerGly 205
Db 541 AGGACATGGGCTATAGAATAATTTTACTCTAGCCAGGAATAGTGGATGACAGCGGA 600
QY 206 SerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAspIleTyrLysLysLeu 225
Db 601 TCCACCAAGCTTTATGAAACTGAACAAGTCCGCAATGTCGATATCTATAAAAACTG 660
QY 226 TyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArgCysIleAlaCysGly 245
Db 661 TACCACAGTGTATGCTGCTTCTTCAAAGCAGTGGTTCCTTACGCTGTATAGCTCGCGG 720
QY 246 ValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGluSerAlaLeuProGly 265
Db 721 GTCAACTTGAACCTCAAGCGCCAGAGCAGGATTGTGGCGCGGAGAGCGGCTCCCGGG 780
QY 266 AlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisValCysGlyGlySerIle 285
Db 781 GCCTGGCCCTGGCAGGTCAAGCTGCAGTCCAGAACGTCACAGTGTGGGAGGCTCCATC 840
QY 286 IleThrProGluTrpIleValThrAlaAlaHisCysValGluLysProLeuAsnAsnPro 305
Db 841 ATCAACCCCGAGTGGATCTGTGACAGCGCCCACTGCGTGGAAANNNNNNNNNNNNN 900
QY 306 TrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGlyAlaGly 325
Db 901 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 960
QY 326 TyrGlnValGluLysValIleSerHisProAsnTyrAspSerLysThrLysAsnAsp 345
Db 961 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1020
QY 346 IleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuValLysProValCys 365
Db 1021 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1080
QY 366 LeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrpIleSerGlyTyrGly 385
Db 1081 CTGCCCAACCCAGGATGATGCTGCAGCAGACAGCTCTGTGGATTTCGGGTGGGGG 1140
QY 386 AlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAlaLysValLeuLeu 405
Db 1141 GCCACCCGAGGAGAAAGGAGACCTCAGAAAGTGTGAACGCTGCCAAGGTGTCTCAT 1200
QY 406 GluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIleThrProAlaMetIle 425
Db 1201 GAGACACAGATGCAACAGCAGATATGTCTATGCAACCTGATCACACAGCATGATC 1260
QY 426 CysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyProLeu 445
Db 1261 TGTCCCGGCTTCTTCAGGGGAAACGTCGATTCTTTCAGGGGTGACAGTGGAGGCGCTCTG 1320

```

```

Qy 446 ValThrSerLysAsnAsnIleTrpThrLeuIleGlyAspThrSerTrpGlySerCys 465
Db 1321 GTCACTTCGAGACAAATATCTGGTGGCTGATAGGGATACAAAGCTGGGGTTCGGCTGT 1380
Qy 466 AlaLysAlaLysThrArgProGlyValTyrglyAsnValMetValPheThrAspTyrIleTyr 485
Db 1381 GCCAAGCTTACAGACAGAGGTGACGGGAATGTGGTATTACAGGACTGGATTAT 1440
Qy 486 ArgGlnMetArg 489
Db 1441 CGACAAATGAGG 1452

RESULT 3
LOCUS AY419483
DEFINITION Mus musculus TMPRSS2 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY419483
VERSION AY419483.1 GI:39775440
KEYWORDS GSS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1446)
Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tannenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Periera,S., Wang,G.; Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
14671302
2 (bases 1 to 1446)
Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tannenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Periera,S., Wang,G.; Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Source 1..1446
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>1446
/gene="TMPRSS2"
/locus_tag="HCM6905"
Gene
ORIGIN
Alignment Scores:
Pred. No.: 6,766-156 Length: 1446
Score: 1770.00 Matches: 330
Percent Similarity: 75.41% Conservative: 35
Best Local Similarity: 68.18% Mismatches: 117
Query Match: 65.15% Indels: 2
DB: 29 Gaps: 2

US-09-615-285b-2 (1-492) x AY419483 (1-1446)
Qy 6 GlySerProAlaIleGlyProTyrTyrglyAsnHisGlyTyrglnProGluAsnPro 25
Db 1 CGGTCACTCCAGGAATCGACCTTGCTATGAGAACCAACCGGCTATCAGTCTGAGCACATC 60
Qy 26 TyrProAlaGlnProThrValValProThrValTyrglyAsnValHisProAlaGlnTyrTyr 45
Db 61 TGTCTCCGAGACCAACGAGTGGCTCCCAATGGTACACTGTATCGAGCCCACTACTAC 120
Qy 46 ProSerProValProGlnTyrAlaProArgValLeuThrGlnAlaSerAsnProValVal 65

```

```

Db 121 CCAATCTCCAGTGCCTCAGATATGCTCCGAGGATTACACGCAAGCCTCAACATCTGTCTC 180
Qy 66 CysThrGlnProLysSerProSerGlyThrValCysThrSerLysThrLysLysAlaLeu 85
Db 191 CACACACATCCCAAGTCC--TCAGGAGCACCCGTCGACCTCAAAGTCTAAGAAATCCTG 237
Qy 86 CysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeuAlaAlaGlyLeuLeu 105
Db 238 TGTTTAGCCCTTGCCCTGCGCACTGCTCTCACGGAGCTGCTGTGGCTGTCTGCTTCTT 297
Qy 106 TrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCysAspSerSerGlyThr 125
Db 298 TGGAGGTTCGGGACAGCACTGTTCTACGTCAGAGTGGAGTGGGTCTTCAGGCACA 357
Qy 126 CysIleAsnProSerAsnTrpCysAspGlyValSerHisCysProGlyGlyGluAspGlu 145
Db 358 TGCATCAGCTCTTCTCTCTGTTGTGACGGGTAGCACATTGTCCCAACGAGAGATGAG 417
Qy 146 AsnArgCysValArgLeuTyrglyProAsnPheIleuGlnValTyrsSerGlnArg 165
Db 418 AACCGTTGTGTTGCTCTACGGACAAAGCTTCATCTCCAGGTTTACTCATCTCAGAG 477
Qy 166 LysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyrglyArgAlaAlaCys 185
Db 478 AAAGCTCGTATCCCGTGTGCCAGATGATTGGAGTCAGAGCTACGGGACAGCATGT 537
Qy 186 ArgAspMetGlyTyrglyAsnAsnPheTyrsSerSerGlnGlyIleValAspSerGly 205
Db 538 AAAGACATGGGATCAAGAACAAATTTTATTCTAGCCAGGGATACACAGACGAGCGG 597
Qy 206 SerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAspIleTyrglyLysLeu 225
Db 598 GCACGAGCTTTATGAGCTGAATGTGAGCTCAGGCAATGTGACCTCTATAAAACTC 657
Qy 226 TyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArgCysIleAlaCysGly 245
Db 658 TACCACAGTACTGTATCCCGCATGGTGTTCCTTTGCGCTGTATAGATGCGGG 717
Qy 246 ValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGluSerAlaLeuProGly 265
Db 718 GTTCGC--TCAGTGMAACGCCAGCAGCAGGATTGTGGTGGATTGAATGCTCACCAGGA 774
Qy 266 AlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisValCysGlyGlySerIle 285
Db 775 GACTGGCCCTGGCAGGTACGCTCAGCTCAGGCGTCCAGCTCTCGGAGGCTCCATC 834
Qy 286 IleThrProGluTrpIleValThrAlaAlaHisCysValGluLysProLeuAsnAsnPro 305
Db 835 ATCACCCCGAGTGGATTGTGACGGCGCCCACTGTGTGGAAGANNNNNNNNNNNNNNN 894
Qy 306 TrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrglyAlaGly 325
Db 895 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 954
Qy 326 TyrGlnValGluLysValIleSerHisProAsnTyrsAspSerLysThrLysAsnAsp 345
Db 955 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1014
Qy 346 IleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuValLysProValCys 365
Db 1015 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1074
Qy 366 LeuProAsnProGlyMetMetLeuGlnProGluLeuCysTrpIleSerGlyTrpGly 385
Db 1075 CTGCGCAACCCAGGCATGATGCTAGACCTAGACAGCAAGATGCTGGATTTCGGGGTGGGG 1134
Qy 386 AlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAlaLysValLeuLeuIle 405
Db 1135 GCCACCTATGAGAAGGAAAGACCTCGAGCTGTGTAAGTCTGCCATGTGTACCTTGATC 1194
Qy 406 GluThrGlnArgCysAsnSerArgTyrglyValTyrsAsnLeuIleThrProAlaMetIle 425

```

```

Db      1195 GAGCCCTCCAAATGTAATAGTAATAATACATACACCAACCTAATACACACGCGCATGATC 1254
Qy      426 CysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyGlyProLeu 445
Db      1255 TGTCCGCGCTTCCCTCCAGGGGTCTGTGCACCTCTTCCAGGGAGACATGGAGGCGCCCTG 1314
Qy      446 ValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSerTrpGlySerGlyCys 465
Db      1315 GTTACTTTGAAGAAATGGGATCTGTGGTGTATTGGGACACAGCTGGGGCTCGGGCTGT 1374
Qy      466 AlalysAlaTyArgProGlyValTyGlyAsnValMetValPheThrAspTrpIleTy 485
Db      1375 GCCAAGGCACATCAGACCTGGAGTATACGGGAACGTGACGGTAATTTACAGATTGGATCTAC 1434
Qy      486 ArgGlnMetArg 489
Db      1435 CAGCAATGAGG 1446

RESULT 4
AY419482
LOCUS
DEFINITION
  Pan troglodytes TMPRSS2 gene, VIRTUAL TRANSCRIPT, partial sequence,
  genomic survey sequence.
ACCESSION
  AY419482
VERSION
  AY419482.1 GI:39775439
KEYWORDS
  GSS.
SOURCE
  Pan troglodytes (chimpanzee)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
  1 (bases 1 to 1452)
AUTHORS
  Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
  Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
  Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
  Adams,M.D. and Cargill,M.
TITLE
  Inferring nonneutral evolution from human-chimp-mouse orthologous
  gene trios
JOURNAL
  Science 302 (5652), 1960-1963 (2003)
PUBMED
  14671302
REFERENCE
  2 (bases 1 to 1452)
  Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
  Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
  Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
  Adams,M.D. and Cargill,M.
  Direct Submission
  Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
  Rockville, MD 20850, USA
  These sequences were made by sequencing genomic exons and ordering
  them based on alignment.
FEATURES
  source
    1..1452
    /organism="Pan troglodytes"
    /mol_type="genomic DNA"
    /db_xref="taxon:9598"
    <1..>1452
    /gene="TMPRSS2"
    /locus_tag="HCM6905"
ORIGIN
Alignment Scores:
Pred. No.:      6,8e-156      Length:      1452
Score:          1770.00      Matches:      337
Percent Similarity: 70.04%      Conservative: 2
Best Local Similarity: 69.63%      Mismatches: 145
Query Match:    65.15%      Indels:      0
DB:             29          Gaps:         0

US-09-615-285B-2 (1-492) x AY419482 (1-1452)
Qy      6 GlySerProAlaIleGlyProTyTrpTyGluAsnHisGlyTyGlnProGluAsnPro 25
Db      1 GGGTCACCAACAGCTATTGGACCTTACTATGAAACCACTGGATACCAACCGAAAAACCCC 60

```

```

Qy      26 TyrProAlaGlnProThrValValProThrValTyGluValHisProAlaGlnTyTrp 45
Db      61 TATCCCGCACAGCCCACTGTGGCCCCCACTGTCTACGAGGTGCATCCGCGCTCAGTACTAC 120
Qy      46 ProSerProValProGlnTrpAlaProArgValLeuThrGlnAlaSerAsnProValVal 65
Db      121 CGGTCCCCCGTCCCGCAGTACGCCCGGAGGTCCTGACGACAGCTTCCAAACCCCGTCCGTC 180
Qy      66 CysThrGlnProLysSerProSerGlyThrValCysThrSerLysThrLysLysAlaLeu 85
Db      181 CGCATGCAGCCCAAAATCCCACTCCGGGACAGTGTGCACCTCAANNNNNNNNNNNNNNNNNN 240
Qy      86 CysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeuAlaGlyLeuLeu 105
Db      241 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 300
Qy      106 TrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCysAspSerSerGlyThr 125
Db      301 NNNNNNNNNNTGGCAGCAAGTCTCCAACTCTGGGATAGAGTGGCACTCTCTCAGGTACC 360
Qy      126 CysIleAsnProSerAsnTrpCysAspGlyValSerHisCysProGlyGlyGluAspGlu 145
Db      361 TGCATCAGCCCTCTAACTGGTGTGATGGCGTGCACCTGCCCCAGCGGGGAGACGAG 420
Qy      146 AsnArgCysValArgLeuTyGlyProAsnPheIleLeuGlnValTyTrpSerGlnArg 165
Db      421 AATCGGTGTGTCTCGCTCTACGACCAAACTTCATCTTCAGTGTACTCATCTCAGAGG 480
Qy      166 LysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyGlyArgAlaAlaCys 185
Db      481 AAGTCTCGCACCTCTGTGCCAAGACGACGTGGAACGAGACTACGGCGGGCGGCGCTGC 540
Qy      186 ArgAspMetGlyTyTrpLysAsnAsnPheTyTrpSerGlnGlyIleValAspSerGly 205
Db      541 AGGACATGGGCTATGAAGTAATATTTTACTTAGCCCAAGGAATAGTGGATGACACGGGA 600
Qy      206 SerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAspIleTyTrpLysLeu 225
Db      601 TCCACCAGCTTTATGAACCTGAACACAGTGGCGGCAATGTGATATCTATAAAAAACTG 660
Qy      226 TyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArgCysIleAlaCysGly 245
Db      661 TACCACAGTATGCTCTGTCTTCAAAGCAGTGTGTTCTTTACGCTGTATAGCTCGGG 720
Qy      246 ValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGluSerAlaLeuProGly 265
Db      721 GTCAACTTGAACCTCAAGCCGCCAGACAGAGATCGTGGTGGCGAGACGCGCTCCCGGG 780
Qy      266 AlaTrpProGlnValSerLeuHisValGlnAsnValHisValCysGlyGlySerIle 285
Db      781 GCCTGGCCCTGGCAGGTACGCTGCACGTCCAGAACGTCCAGTGTGGGAGGCTCCATC 840
Qy      286 IleThrProGluTrpIleValThrAlaAlaHisCysValGluLysProLeuAsnAsnPro 305
Db      841 ATCACCACCCGAGTGGATCGTGACAGCGCCCACTGGGTGGAAANNNNNNNNNNNNNNNN 900
Qy      306 TrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyGlyAlaGly 325
Db      901 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 960
Qy      326 TyrGlnValGluLysValIleSerHisProAsnTyArgSerLysThrLysAsnAsp 345
Db      961 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1020
Qy      346 IleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuValLysProValCys 365
Db      1021 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1080
Qy      366 LeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrpIleSerGlyTrpGly 385
Db      1081 CTGCCCAACCCAGGATGATGCTGGAGCCAGACAGCTCTGCTGATTTCCGGGTGGGG 1140
Qy      386 AlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAlaLysValLeuLeu 405

```


AUTHORS Poustka, A., Albert, R., Moosmayer, P., Schupp, I., Wellenreuther, R.,
Mewes, H.W., Weil, B., Amid, C., Osanger, A., Fobo, G., Han, M. and
Wiemann, S.
TITLE EST (Foustka, A., Albert, R., Moosmayer, P., Schupp, I.,
Wellenreuther, R., et al.)
JOURNAL Unpublished (2003)
COMMENT Contact: MIPS

MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.

No s1 sequence available.
This clone (DKFZp686C09277) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
Location/Qualifiers
1..781
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686C09277"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplex2; Site_1: SfIIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN

Alignment Scores:
Pred. No.: 2,46e-102 Length: 781
Score: 1198.00 Matches: 215
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.54% Mismatches: 0
Query Match: 44.0% Indels: 0
DB: 13 Gaps: 0

US-09-615-285B-2 (1-492) x BX508967 (1-781)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 132 ATGGCTTTGAATCAGGCTCACCACCGACTATGGACCTTACTATGAAACCATGGATAC 191
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
DB 192 CAACCGAAACCCCTATCCCGACAGCCCACTGTGGTCCCACTGTCTACGAGTGCAT 251
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 252 CCGGCTCAGTACTACCCGTCCTCCCGGTCCTCCAGTACGCCCCGAGGCTCTGACGAGGCT 311
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 312 TCNACCCCGTGTCTGTCAGCGAGCCCAATCCCATCCGGACAGTGTGCACCTCAAAG 371
QY 81 ThrLysLeuAlaLeuAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
DB 372 ACTAAGAAGCACTGTGTCATCACTTGACCTGGGACCTTCTCTGTGGAGCTCGGTG 431
QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB 432 CGCGCTGGCTACTCTGGAAGTTCATGGGAGCAAGTGTCCCACTCTGGGATAGATGC 491
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
DB 492 GACTCCTCAGGTACCTGCATCAACCCCTCTAATGCTGTGATGGGTGTGCACACGCCCC 551
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
DB 552 GGCAGGAGGAGGAGGAGTGGTGTGTGGCTTACGGACCAACTTCTCATCTTCTCAGATG 611

QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGlnuSerTyr 180
DB 612 TACTCATCTCAGAGGAAGTCTCTGGCACCCCTGTGTGCCAAGACGACTGGAAACGAACTAC 671
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
DB 672 GGCAGGCGCGCTCTGAGGACATGGCTATAGATAATTTTACTCTAGCCAGGAATA 731

QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAla 216

DB 732 GTGGATGACAGCGGATCCACCGACTTATGAAACTGAACACCAAGTGCC 779

RESULT 7

BX479625

LOCUS

DEFINITION

DKFZp686E22213_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BX479625 603 bp mRNA linear EST 04-SEP-2003
DKFZp686E22213_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DKFZp686E22213_5', mRNA sequence.

ACCSSION BX479625

VERSION BX479625.1 GI:31915563

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 603)

Boecker, H., Boecker, M., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,

Fobo, G., Han, M. and Wiemann, S.

EST (Boecker, H., Boecker, M., Mewes, H.W., Weil, B., Amid, C., et al.)

Unpublished (2003)

Contact: MIPS

MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by GBR (National Research Centre for Biotechnology Ltd.,

Braunschweig/Germany) within the cDNA sequencing consortium of the

German Genome Project.

No s1 sequence available.

This clone (DKFZp686E22213) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers

1..603

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZp686E22213"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="686 (synonym: hlcc3)"

/note="Vector: pTriplex2; Site_1: SfIIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN

Alignment Scores:
Pred. No.: 2,88e-93 Length: 603
Score: 1100.00 Matches: 200
Percent Similarity: 99.50% Conservative: 0
Best Local Similarity: 99.50% Mismatches: 1
Query Match: 40.49% Indels: 0
DB: 13 Gaps: 0

US-09-615-285B-2 (1-492) x BX479625 (1-603)

QY 145 GluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnValTyrSerSerGln 164
DB 1 GAGAAATCGGTGTGTTCGCCCTCAGGACTAACTTATCTCCTCAGGTGTACTCTCTCAG 60
QY 165 ArgLysSerTrpHisProValCysGlnAspAspTrpAsnGlnuSerTyrGlyArgAlaAla 184
DB 61 AGGAAGTCTCTGGCACCCCTGTGTGCCAAGACGACTGGAAACGAGAACTACGGCGCGCGGCC 120

QY 185 CysArgAspMetGlyTyrIysAsnAsnPheTyrSerSerGlnGlyIleValAspAspSer 204
 DB 121 TGCAGGAGCAATGGCTATAGATAATATTTTACTCTAGCCAAAGGAATAGTGATGACAGC 180
 QY 205 GlySerThrSerPheMetIysLeuAsnThrSerAlaGlyAsnValAspIleTyrIysLys 224
 DB 181 GGATCCACCACTTTATGAACCTGAACCAAGTCCGCAATGTCTGATATCTATAAAAA 240
 QY 225 LeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArgCysIleAlaCys 244
 DB 241 CTGTACCAACAGTATGCTGCTTTCTCAAGCAGTGGTTCTTTACGCTGTATACCTGCG 300
 QY 245 GlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlySerAlaLeuPro 264
 DB 301 GGGGTCAACTGAACCTCAAGCCGCCAGACGAGCATTTGTGGCGGCGAGAGCGCTCCCG 360
 QY 265 GlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisValCysGlyGlySer 284
 DB 361 GGGGCTGGCCCTGGCAGGTACCTGACGTCCAGACGTCCAGTGTGGAGGCTCC 420
 QY 285 IleIleThrProGluTyrIleValThrAlaAlaHisCysValGluGlyProLeuAsnAsn 304
 DB 421 ATCATCAACCCCGAGTGGATCGTGACAGCGCCCACTGGTGGAAAAACCTCTTAACAA 480
 QY 305 ProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGlyAla 324
 DB 481 CCATGGCATTTGGACGCAATTTGGGGGATTTTGAGCAATCTTTCAATGTTCTATGGAGCC 540
 QY 325 GlyTyrGlnValGluIysValIleSerHisProAsnTyrAspSerLysThrIysAsnAsn 344
 DB 541 GGATCAACCAAGTAGAAAAAGTATTTCTCATCAAAATTTAGCTCCCAAGACCAAGAACAA 600
 QY 345 Asp 345
 DB 601 GAC 603

RESULT 8
 BI414727
 LOCUS 602991112F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5147162 5',
 DEFINITION mRNA sequence.
 ACCESSION BI414727
 VERSION BI414727.1 GI:15175650
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 921)
 TITLE NIH-MGC http://mgc.nci.nih.gov/.
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLML1363 row: 9 column: 03
 High quality sequence start: 29
 High quality sequence stop: 873.
 Location/Qualifiers
 1..921
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="Czech II"
 /db_xref="taxon:10090"
 /clone="IMAGE:5147162"

/tissue type="pooled lung tumors"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NCI_CGAP_Lu33"
 /modified linker; Vector: pVT73D-Pac (Pharmacia) with a
 modified polylinker; Site: 1: NotI; Site: 2: EcoRI; 1st
 strand cDNA was prepared from mRNA obtained from pooled
 lung tumors with a Not I - oligo(dT) primer [5',
 TGTTCACCAATCTGAAGTGGAGCGCGCTCTGTTTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified pVT73 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN
 Alignment Scores:
 Pred. No.: 1,53e-92 Length: 921
 Score: 1095.50 Matches: 206
 Percent Similarity: 82.93% Conservative: 32
 Best Local Similarity: 71.78% Mismatches: 47
 Query Match: 12 Indels: 3
 DB: 1 Gaps: 1

US-09-615-285B-2 (1-492) x BI414727 (1-921)

QY 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 DB 65 ATGGATTTGAATCTAGGCTCACTCCAGGATCCGACTTGCATATGAGACCCAGGAT 124
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 DB 125 CAGTCTGAGCACATCTGTCTCCGAGACCCAGCATGGCTCCCAATGGCTCAACTTCTAT 184
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 DB 185 CCAGCCCGAGTACTACCCATCTCCAGTGCCTCAGTATGCTCCGAGGATTACACGCAAGCC 244
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 DB 245 TCAACATCTGTATCATCCACACATCCCAAGTCC---TCAGGAGCACCGCTGCACCTCAAG 301
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 DB 302 TCTAAGAAATCGTGTGTAGCCCTTGGCCCTGGGCACTGTCTCCACGGGAGCTGTCTGTG 361
 QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 DB 362 GTGTGTCTGTCTGTGTGGAGGTTCTGGACAGCAACTGTCTTCTACGCTCGAGATGGAGTGT 421
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 DB 422 GGGTCTTCAGGCACATGATCAGCTCTCTCTCTGGTGTGACGGGTAGCACATTGTGCC 481
 QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 DB 482 AACGGAGAGATGAGAACCGTTGTGTCTCTACGGACAAAGCTTCATCTCCAGGTT 541
 QY 161 TyrSerSerGlnArgLysSerTyrPheIleProValCysGlnAspAspTyrAsnGluAsnTyr 180
 DB 542 TACTCATCTCAGAGGAAAGCCTGGTATCCCGTGTCCAGGATGATTTGAGTAGAGAGTAC 601
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrIleAsnAsnPheTyrSerSerGlnGlyIle 200
 DB 602 GGGAGAGCAGCATGTAAAGACATGGGATACAGAACAAATTTTATTCTACCAAGGAGATA 661
 QY 201 ValAspAspSerGlySerThrSerPheMetIysLeuAsnThrSerAlaGlyAsnValAsp 220
 DB 662 CCAGACCAAGACGGGGCAACGAGCTTTATGAAGCTGAATGTGAGCTCAGGCAATGTTGAC 721
 QY 221 IleTyrLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 DB 722 CTCTATAAAACACTCTACCAAGTGAATTCAGTTTCATCCCGCATGTTCTTTTGGCC 781

QY 241 CyslleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArglleValGlyGlu 260
 782 TGATAGAAATCGGGGGTTCCTCAGCA-AACCGCAAGACAGGATGTGGGGGATG 840
 QY 261 SerAlaLeuProGlyValAlaTrpProTrpGlnValSerLeuHisValGln-AsnValHisVa 280
 841 AATGCTCACCAGGAGAGTGGCTGGCAGGTGAGCTTGACAGTCCCAATGGCGTCACCGT 900
 QY 280 lCysGlyGlySerllelle 286
 Db 901 CTGGCGAGGCTCCATCATC 919

RESULT 9
 CD724002
 LOCUS
 DEFINITION
 oJ29a11.y1 Human lacrimal gland, unamplified: oJ Homo sapiens cDNA
 clone oJ29a11 5', mRNA sequence.

ACCESSION
 CD724002
 VERSION
 CD724002.1 GI:32274856
 KEYWORDS
 EST.
 SOURCE
 Homo sapiens (human)

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 (bases 1 to 604)
 Dickinson,D., Laurie,G. and Wistow,G.

AUTHORS
 Expressed sequence tag analysis of human lacrimal gland

TITLE
 Unpublished (2002)

JOURNAL
 Contact: Wistow G

COMMENT
 Section on Molecular Structure and Function

National Eye Institute

6/331, NIH, Bethesda, MD 20892-2740, USA

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Place: 29 row: a column: 11

Seq primer: M13R1 reverse primer (ABI).

Location/Qualifiers

1..604

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="oJ29a11"

/tissue_type="lacrimal gland"

/dev_stages="Adult"

/lab_host="EMDH10B"

/clone_lib="Human lacrimal gland, unamplified: oJ"

/note="Organ: Eye; Vector: pCMVSPORT6; RNA was extracted

from 2 human lacrimal glands. A directionally cloned cDNA

library in the pCMVSPORT6 vector(Life Technologies) was

constructed at Bioserve Biotechnology (Laurel MD)

essentially following the protocols of the Superscript

Plasmid System full details of which are contained in the

manufacturer's instruction manual

(http://www.lifetech.com/). First strand synthesis was

carried out using a Not I primer-adaptor

[5'-PGACTAGTCTAGATCGGAGCGGCC(T)15-3']. EST analysis

was performed on the unamplified library at the NIH

Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:

Pred. No.: 3,92e-92 Length: 604

Score: 1088.00 Matches: 200

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 40.04% Indels: 0

DB: 14 Gaps: 0

US-09-615-285b-2 (1-492) x CD724002 (1-604)

QY 183 AlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlylleValasp 202

Db 3 GCGGCTCGCAGGACATCGGGCTATAGAATAATTTTACTCTAGCAAGAAATAGTGGAT 62
 QY 203 AspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAspIleTyr 222
 Db 63 GACAGCGGATCCACAGCTTTTGAACATGAACAAGTGGCGGCAATTCGATATCTAT 122
 QY 223 LysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArgCysIle 242
 Db 123 AAAAAACTGTACCACAGTGTCTCTCTCAAAAGCAGTGTCTTTTACGCTGATA 182
 QY 243 AlaCysGlyValAsnLeuAsnSerSerArgGlnSerArglleValGlyGlyGluSerAla 262
 Db 183 GCCTGGGGGTCAACTTGAACCTCAAGCCGACAGAGGATCTGTGGCGGCGAGACGCG 242
 QY 263 LeuProGlyValAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisValCysGly 282
 Db 243 CTCCCGGGGGCTGGCCCTGGCAGGTTCAGCTCCAGAACCTCCACGTTGCGGA 302
 QY 283 GlySerllelleThrProGlnTrpIleValThrAlaAlaHisCysValGluLysProLeu 302
 Db 303 GGCTCCATCATCACCCCGAGTGGATCGTGACAGCGCCGACCTGCGTGGAAAAACCTCTT 362
 QY 303 AsnAsnProTrpHisTrpThrAlaPheAlaGlylleLeuArgGlnSerPheMetPheTyr 322
 Db 363 AACAAATCCATGGCATTTGACGGCATTTGCGGGGATTTTGAGACAATCTTTCATGTTCTAT 422
 QY 323 GlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerlysthrLys 342
 Db 423 GGAGCGGATACCAAGTAGAAAAAGTGAATTTCTCATCAATATGACATCCAAAGACCAAG 482
 QY 343 AsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuValLys 362
 Db 483 AACAAATGACATTCGCTGATGAAGCTGCAGAGCCTTGACITTCACGACCTTAGTAAA 542
 QY 363 ProValCysLeuProAsnProGlyMetMetLeuGlnProGlnLeuCysTyrPheSer 382
 Db 543 CCAGTGTGCTGCCCCAACCCAGGCATGCTGCTGAGCCAGACAGCTGCTGCGATTTC 602

RESULT 10

CF147359

LOCUS

DEFINITION

UI-HF-CBO-atj-b-03-0-UI.r1 NIH_MGC_210 Homo sapiens cDNA clone

IMAGE:30571118 5', mRNA sequence.

CF147359

ACCESSION

CF147359.1 GI:33262803

KEYWORDS

EST.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 678)

AUTHORS

Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL

97044477

MEDLINE

8889548

PUBMED

COMMENT

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Tim Ratliff

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

http://genome.uiowa.edu/distribution/humanfl.html

Seq primer: pYX-5.

Location/Qualifiers

1..678

FEATURES

source

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30571118"
 /tissue_type="CNCAP (3) T-225 cell line"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH MGC 210"
 /notes="Organ: Prostate; Vector: pRTT3 Pac; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pRTT3 Pac vector. The library tag sequence located between the Not I site and the polyA tail is CCCAC. Tissue was provided by Tim Ratliff."

ORIGIN

Alignment Scores:

Pred. No.: 1.13e-91 Length: 678
 Score: 1084.00 Matches: 196
 Percent Similarity: 99.49% Conservative: 0
 Best Local Similarity: 99.49% Mismatches: 1
 Query Match: 39.90% Indels: 1
 DS: 14 Gaps: 0

US-09-615-285B-2 (1-492) x CF147359 (1-678)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Db 88 ATGGCTTTGAACCTCAGGTCACACAGCATTTGACCTTACTATGAACCATGGATAC 147
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 Db 148 CAACCGGAACACCCCTATCCACAGCCACTGGTGGTCCCATCTGTACAGGTGCAT 207
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 208 CGGCTCAGTACTACCCGTCCTCCGTCAGTACGCCCCAGGCGGCTCGACAGCGCT 267
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 268 TCACACCCCGTCTGTCAGCAGGCGCCAAATCCCATCCGAGCAGTGTGCACCTCAAG 327
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 Db 328 ACTAAGAAAGCACTGTGCATCACTTGCCTGGGACCTTCTCGTGGGAGCTGGCTG 387
 QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 388 GCGCTGGCTACTCTGGAAGTTCATGGGACCAAGTGTCTCAACTCTGGGATAGATGC 447
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 Db 448 GACTCCTCAGGTACCTGCATCAACCCCTTAACCTGTGTGATGGGTGTGCACACTGCC 507
 QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 508 GCGGGGAGCAGCAATCGGTGTCTCGCTCTACGGACCAAACTTCATCTCTCAAGT 567
 QY 161 TyrSerSerGlnArgLysSerThrPheSerProValCysGlnAspAspTyrAsnGluAsnTyr 180
 Db 568 TACTCATCTCAGAGGAGTCTGGCACCCCTGTGTGCCAAGACGACTGGACGAGACTAC 627
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSer 197
 Db 628 GGGCGGGCGGCTCGAGGACATGGGCTATAAGATAT-ATTTACTCTAGC 677

RESULT 11

BM747154

LOCUS

DEFINITION

K-RST0021659 S6SNU620 Homo sapiens cDNA clone S6SNU620-20-F02 5',

mRNA sequence.

BM747154

VERSION

BM747154.1 GI:19076749

EST.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 591)

Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,

Oh,K.J., Cheong,J.E., Schn,H.Y., Kim,J.M., Park,H.S., Kim,S. and

Kim,Y.S.

21C Frontier Korean EST Project 2001

Unpublished (2002)

Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Soeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 20 row: F column: 02

High quality sequence stop: 591.

Location/Qualifiers

1..591

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="S6SNU620-20-F02"

/sex="F"

/tissue_type="Ascites"

/cell_type="Scattering floating"

/cell_line="SNU-820"

/lab_host="Top10F"

/clone_lib="S6SNU620"

/note="Organ: Stomach; Vector: pcNS; Site 1: EcoRI;

Site 2: NotI; The poly (A) + RNA was dephosphorylated with

bacterial alkaline phosphatase (BAP) and then decapped

with tabacco acid pyrophosphatase (TAP). The decapped

inact mRNA was ligated with DNA-RNA linker including EcoR

I site by treatment of T4 RNA ligase and the first strand

cDNA was synthesized from oligo dt-selected mRNA by

priming with dt-tailed vector. The dt-tailed vector was

adjusted to have about 60nt. The cDNA vector was

circularized with E. coli DNA ligase after digestion of

EcoRI which site is also included in vector. An RNA strand

converted to a DNA strand by Okayama-Berg method. The

obtained cDNA vectors were used for transformation of

competent cells E. coli Top10F' by electroporation method.

The cDNA libraries constructed by this method are

full-length enriched cDNA library."

ORIGIN

Alignment Scores:

Pred. No.: 4.51e-90 Length: 591
 Score: 1066.00 Matches: 196
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 39.23% Indels: 0
 DS: 12 Gaps: 0

US-09-615-285B-2 (1-492) x BM747154 (1-591)

QY 224 LysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArgCysIleAla 243
 Db 2 AAACCTGACCACTGATGCTGTTCTTCAAGCAGTGGTCTTCTTACGCTGTATAGCC 61
 QY 244 CysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGluSerAlaLeu 263
 Db 62 TCGCGGGTCAACTTGAATCAAGCCGACAGCAGGATTGTGGCGCGCGCGCGCTC 121
 QY 264 ProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisValCysGlyGly 283


```

Db 122 CCGGGGGCTGGCCCTGGCAGCTGCGCTCCAGAGCTCCACGCTGTGCGGAGGC 181
QY 284 SerileileThrProGluThrPileValThrAlaAlaHisCysValGluLysProLeuAen 303
Db 182 TCCATCATCACCCCGAGTGGATCGTGACAGCGCCGACCTGCGTGAAAAACCTCTTAAC 241
QY 304 AsnProThrHisThrAlaPhaAlaGlyLeuArgGlnSerPheMetPheTyrGly 323
Db 242 AATCATGGCATTGGACGGCAATTTGCGGGGATTTGAGACAATCTTTTCATGTTCTATGGA 301
QY 324 AlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLysThrLysAen 343
Db 302 GCGGATACCAAGTAGAAAAAGTATTTCTCATCCAAATATGACTCCAGACCAAGAAC 361
QY 344 AsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuValLysPro 363
Db 362 AATGACATTGGCTGGATGAAGCTGACAGAGCCCTGACTTTCAACCACTAGTGAACA 421
QY 364 ValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrpIleSerGly 383
Db 422 GTGTGTCTGCCAACCCAGCATGATGCTGCAGCCAGACAGCTCTGCTGGATTTCCGGG 481
QY 384 TrpGlyAlaThrGluGlnLysGlyLysThrSerGluValLeuAsnAlaLysValLeu 403
Db 482 TGGGGGGCCACCGAGGAGAAAGGAAAGCTCAGAAAGTGTGAACGCTGCCAAGGTGCTT 541
QY 404 LeuileGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeu 419
Db 542 CTCATTGACACAGAGATGCACAGAGATATGTTATGACACACCTG 589

```

```

RESULT 12
B1454294
LOCUS 603171527F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5250822 5',
DEFINITION mRNA sequence.
ACCESSION B1454294
VERSION B1454294.1 GI:15244950
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

```

```

REFERENCE 1
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLES NIH-MGC http://mgs.nci.nih.gov/.
NATIONAL INSTITUTES OF HEALTH, Mammalian Gene Collection (MGC)
UNPUBLISHED (1999)
CONTACT: Robert Strausberg, Ph.D.
EMAIL: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11633 row: f column: 07
High quality sequence stop: 736.
Location/Qualifiers
1..736
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5250822"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam5"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site: 1; Salt;
Site: 2; Not; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,

```

FEATURES

```

Source
1..736
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5250822"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam5"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site: 1; Salt;
Site: 2; Not; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,

```

```

RESULT 13
CB989588
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS

```

```

CB989588 767 bp mRNA linear EST 01-MAY-2003
AGENCOURT 13894812 NIH MGC_147 Homo sapiens cDNA clone
IMAGE:30347890 5', mRNA sequence.
CB989588
VERSION
CB989588.1 GI:30284108
EST.

```

```

ORIGIN
Alignment Scores: 4.48e-88 Length: 736
Pred. No.: 1046.50 Matches: 193
Score: 87.45% Conservative: 23
Percent Similarity: 78.14% Mismatches: 28
Best Local Similarity: 38.52% Indels: 3
Query Match: 12 Gaps: 1
DB:
US-09-615-285B-2 (1-492) x B1454294 (1-736)
QY 98 AlaAlaLeuAlaGlyLeuLeuThrLysPheMetGlySerLysCysSerAsnSerGly 117
Db 2 GCTGCTGTGGCTGCTCTGCTTTGGAGGTTCTGGACAGCAACTGTTCTACGCTCTAG 61
QY 118 IleGluCysAspSerSerGlyThrCysIleLeuAsnProSerAsnTyrCysAspGlyValSer 137
Db 62 ATGAGTGTGGTCTTCAGGCACATGCATCAGCTCTCTCTCTGCTGTGACGGGTAGCA 121
QY 138 HisCysProGlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPhelle 157
Db 122 CATGTGCCCAACGAGAGATGAGAACCGTTGTGTCTCTACGGACAAAGCTTCATC 181
QY 158 LeuGlnValTyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsn 177
Db 182 CTCAGGTTTACTCATCTCAGAGGAAGCTCGTATCCGCTGTCAGAGATGATTGGAGT 241
QY 178 GluAsnTyrGlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPhetTrpSer 197
Db 242 GAGAGCTCGGGAGGAGCAGCATGTAAAGACATGGGATACAGAACAAATTTTATTCTAGC 301
QY 198 GlnGlyIleValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGly 217
Db 302 CRAAGGATACAGAGACAGAGCGGGCAACGACCTTTATGAGCTGAATGTGAGCTCAGGC 361
QY 218 AsnValAspIleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValVal 237
Db 362 AATGTTGAGCTCTATAAAAACTCTACCACAGTACTCATGTTCACTCCGATGGTGGT 421
QY 238 SerLeuArgCysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleVal 257
Db 422 TCTTTGCGCTGTATAGATCGGGTTCGC--TCAGTGAACACGCGCAGAGGATTGTG 478
QY 258 GlyGlyGluSerAlaLeuProGlyAlaThrProTyrGlnValSerLeuHisValGlnAsn 277
Db 479 GGTGGATTGAATGCTCTACAGAGAGACTGGCCCTGCGAGGTGAGCTGCAGCTCAAGGC 538
QY 278 ValHisValCysGlyGlySerIleIleThrProGluThrPileValThrAlaAlaHisCys 297
Db 539 GTCCACGCTCGGGAGGCTCCATCATCACCCTCCAGGTGATGATTGTGACGGCGCCACTGT 598
QY 298 ValGluLysProLeuAsnAspProTrpHis--TrpThrAlaPheAlaGlyIleLeuArgG 317
Db 599 GTGGAAAGAACCCCTCAGCGGGCCCGAGGTACTGTGACGGCAATTCGGGAATTCGAGACA 658
QY 317 nSerPheMetPheTyrGlyAlaGlyTyrGlnValGluLys--ValIleSerHisProAsn 337
Db 659 GTCTCTCATGTTCTATGGAAGTAGACACACAGGTAGAAACAGTAAATTTCCCATCAATT 718
QY 337 YrAspSerLysThrLys 342
Db 719 ACGACTCTAAGACCAAG 735

```

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 767)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. Stefan Hansson
 CDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
 and advice from Piero Carninci (RIKEN)
 CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDAM389 row: 1 column: 11
 High quality sequence stop: 676.

FEATURES
 source
 1..767
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30347890"
 /tissue_type="Human Placenta"
 /lab_host="DH10B TconA"
 /clone_lib="NIH_MGC_147"
 /note="Organ: Placenta; Vector: pBluescriptR; Site 1:
 all-XhoI; Site 2: BamH; Oligo-dr primed using primer
 5'-TTTTTTTTTTTTTTTNN-3', size-selected for average
 insert size 2.3 kb and normalized to ROT 5. This is a
 primary library enriched for full-length clones and
 constructed using the Cap-trapper method (Carninci, in
 preparation). Library constructed by M. Brownstein, in
 (NHGRI/NHGRI, National Institutes of Health). Note: This is
 a NIH_MGC library."

ORIGIN

Alignment Scores:
 Pred. No.: 1,97e-87 Length: 767
 Score: 1040.00 Matches: 208
 Percent Similarity: 79.55% Conservative: 2
 Best Local Similarity: 78.79% Mismatches: 8
 Query Match: 38.28% Indels: 46
 DB: 14 Gaps: 4

US-09-615-285B-2 (1-492) x CB989588 (1-767)

QY 1 MetAlaLeuAsnSerGlySerProAlaLeuGlyProTyrTyrGluAsnHisGlyTyr 20
 Db ATGGCTTTGAATCAGGTCACACACAGCTATTGACCTTACTATGAAACCATGGATAC 155
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 Db 156 CAAACCGGAAACCCCTATCCGACAGCCACTGTGTGTCCTCCCACTGTCTACAGGTCAT 215
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 216 CCGGCTCATATACCCGTCCTCCCTGCGCCAGTACGCCCCGAGGGTCTTGACACAGCT 275
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 276 TCCACACCGCTGCTGTGACGAGCCCAATCCCATCCGAGCAGTGTGCACCTCAAG 335
 QY 81 ThrLysLysAlaLeuCysThrLeuThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 Db 336 ACTAAGAAGAGACTGTGCATACCTTGACCTTGGGAGCTTCTCTCGTGGGAGCTGCCCTG 395
 QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 396 GCGCTGGCTACTCTGGAAGTTC----- 419

QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 Db 419 ----- 419
 QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 420 -----ATTGCGCTCTACGACCAACTTTCATCCCTTCAGGTG 455
 QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 Db 456 TACTCATCTCAGAGGAGTCTCTGGACCTGTGCCAAGACGACTGAGACGAACTAC 515
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
 Db 516 GGGCGGGGGCTGCGAGGACATGGCTATAGAAATATTTTACTCTAGCCAGGAATA 575
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 576 GTGGATGACAGGGGATCCACGCTTTATGAACTGACACAGTGGCGGCAATGTGAT 635
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuAr 240
 Db 636 ATCTATAAAAACTGTACACAGTGTCTTCTTCAAAAAGCAGTGGTTCCTTTTAC 695
 QY 240 GCysCys---AlaCys-GlyValAsn---LeuAsnSerSer-----ArgGlnSerArgI 256
 Db 696 NCGTGTATAAGCTTGGGGGGTCCAACTTGAACCTCAAGCCCGCCAGGAGCCAGGAA 755
 QY 256 leValGly 258
 Db 756 TCGTTGGG 763
 RESULT 14
 CF147031
 LOCUS
 DEFINITION UI-HF-C80-atg-a-09-0-UI-r1 NIH_MGC_210 Homo sapiens CDNA clone
 IMAGE:30570824 5', mRNA sequence.
 CF147031
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 644)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Tim Ratliff
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/humanfl.html>
 Seq primer: pyx-5.

FEATURES
 source

Location/Qualifiers
 1..644
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30570824"
 /tissue_type="CNCAP (3) T-225 cell line"

/lab host="DH10B (T1 phage resistant)"
/clone lib="NIH MGC 210"
/note="Organ: Prostate; Vector: pT73 Pac; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with Not I and then cloned
directionally into pT73 Pac vector. The library tag
sequence located between the Not I site and the polyA tail
is CCCAC. Tissue was provided by Tim Ratliff."

ORIGIN

Alignment Scores:
Pred. No.: 2,01e-86 Length: 644
Score: 1029.00 Matches: 185
Percent Similarity: 99.46% Conservative: 0
Best Local Similarity: 99.46% Mismatches: 0
Query Match: 37.84% Indels: 1
DB: 14 Gaps: 0

US-09-615-285B-2 (1-492) x CF147031 (1-644)

Qy 1 MetAlaLeuAenSerGlySerProProAlaIleGlyProTyTyTyGluAenHisGlyTyr 20
Db 88 ATGGCTTTGAACACTCAGGTCACACCCAGCTATTGGACCTTACTATGAAACCATGGATAC 147

Qy 21 GlnPro-GluAenProTyTyTyProAlaGlnProThrValProThrValTyTyGluValHi 40
Db 148 CAACCGGAAACCCCTATCCCGACACGCCACCTGTGGTCCCCACTCTCTACGAGGTGCA 207

Qy 40 sProAlaGlnTyTyTyProSerProValProGlnTyTyAlaProArgValLeuThrGlnAl 60
Db 208 TCCGCTCAGTACTACCGCTCCCGCTGCCCGACGACGCGGAGGTCTCTGACGACGCG 267

Qy 60 aSerAenProValValCysThrGlnProThrValProThrValCysThrSerLy 80
Db 268 TTCCAACCCCGTGGTCTGACGACGACGACCAATCCCATCCGGAGAGTGTCACCTCAA 327

Qy 80 sThrIlysAlaLeuCysIleThrLeuThrLeuGlyThrPhetLeuValGlyAlaAlaLe 100
Db 328 GACTAAGAAAGCACTGTGCATCACCTTGACCTGGGACCTCTCTCGTGGAGTGGGCT 387

Qy 100 uAlaAlaGlyLeuLeuTyTyTyPheMetGlySerIysCysSerAenSerGlyIleGluCy 120
Db 388 GCGCGCTGGCTACTCTGGAAGTTTCATGGGACGACGACGACGACGACGACGACGACG 447

Qy 120 aAspSerSerGlyThrCysIleAenProSerAenTrpCysAspGlyValSerHisCysPr 140
Db 448 CGACTCTCAGGTACTCTGATCAACCCCTCTAAGTGGTGTGATGGCGGTGCACACTGCC 507

Qy 140 oGlyGlyGluAspGluAenArgCysValArgLeuTyTyGlyProAenPheIleLeuGlnVa 160
Db 508 CGCGGGGAGGACGAGAGTCGGTGTGTTCGCTCTACGACGACCAACTTCATCTTCAGGT 567

Qy 160 lTySerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTy 180
Db 568 GTACTCATCTCAGAGAGATCTGACACCCCTGTGTGCCACGACGACGACGACGACGAC 627

Qy 180 rGlyArgAlaAlaCys 185
Db 628 CGGGCGGGCGGCTGC 643

RESULT 15
CF147378
LOCUS
DEFINITION
IMAGE:30571142 5', mRNA sequence.
ACCESSION
CF147378
VERSION
CF147378.1 GI:33262822
KEYWORDS
EST.

SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 622)
TITLE Ronaldo, M.P., Lennon, G. and Soares, M.B.
JOURNAL Normalization and subtraction: two approaches to facilitate gene
MEDLINE discovery
PUBMED Genome Res. 6 (9), 791-806 (1996)
COMMENT 97044477
8889548
Contact: Soares, MB
Coordinated laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Tim Ratliff
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/humanf1.html
Seq primer: pYX-5.

FEATURES
Location/Qualifiers
1..622
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30571142"
/issue_type="CNCAP(3)T-225 cell line"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH MGC 210"
/note="Organ: Prostate; Vector: pT73 Pac; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with Not I and then cloned
directionally into pT73 Pac vector. The library tag
sequence located between the Not I site and the polyA tail
is CCCAC. Tissue was provided by Tim Ratliff."

ORIGIN

Alignment Scores:
Pred. No.: 8,35e-84 Length: 622
Score: 1000.00 Matches: 178
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.81% Indels: 0
DB: 14 Gaps: 0

US-09-615-285B-2 (1-492) x CF147378 (1-622)

Qy 1 MetAlaLeuAenSerGlySerProProAlaIleGlyProTyTyTyGluAenHisGlyTyr 20
Db 88 ATGGCTTTGAACACTCAGGTCACACCCAGCTATTGGACCTTACTATGAAACCATGGATAC 147

Qy 21 GlnProGluAenProTyTyTyProAlaGlnProThrValProThrValTyTyGluValHis 40
Db 148 CAACCGGAAACCCCTATCCCGACACGCCACCTGTGGTCCCCACTCTCTACGAGGTGCA 207

Qy 41 ProAlaGlnTyTyTyProSerProValProGlnTyTyAlaProArgValLeuThrGlnAla 60
Db 208 CCGGCTCAGTACTACCGCTCCCGCTGCCCGACGACGACGACGACGACGACGACGAC 267

Qy 61 SerAenProValValCysThrGlnProThrValProThrValCysThrSerLy 80
Db 268 TCCAACCCCGTGGTCTGACGACGACGACCAATCCCATCCGGAGAGTGTCACCTCAA 327

QY	81	ThrLysLysAlaLeuCysIleThrIleuThrIleuGlyThrPheLeuValGlyAlaAlaLeu	100
Db	328	ACTAAGAAAGCACTGTGCATCACTTGACCTGGGGACCTTCCTCGTGGGAGCTGGCGTG	387
QY	101	AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys	120
Db	388	GCCGCTGGGCTACTCTGGAGTTCATGGGCGAGAGTGCTCCAACTCTGGGATAGATGC	447
QY	121	AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro	140
Db	448	GACTCCTCAGGTACCTGCATCAACCCCTTAACCTGGTGTGATGGCGTGTACACACTGCCCC	507
QY	141	GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal	160
Db	508	GGCGGGGAGGACGAGATCGGTGTGTTCGCCCTCTACGGACCAAACTTCATCCTTCAGGTG	567
QY	161	TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGlu	178
Db	568	TACTCATCTCAGAGGAAGTCTGGCACCCCTGTGTGCCAAGACGACTGGAACGAG	621

Search completed: June 1, 2004, 17:39:44
Job time : 3794 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 14:07:53 ; Search time 26 Seconds
(without alignments)
985.328 Million cell updates/sec

Title: US-09-615-285B-2

Perfect score: 2717

Sequence: 1 MALNSGSPPAIGPYENHGY.....VYGNVMVFTDIYQMRADG 492

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2714	99.9	492	1 TMS2 HUMAN	O15393 homo sapien
2	2117	77.9	490	1 TMS2 MOUSE	O91088 mus musculus
3	880	32.4	453	1 TMS3 MOUSE	Q81110 mus musculus
4	875.5	32.2	454	1 TMS3 HUMAN	P57727 homo sapien
5	680	25.0	435	1 TMS4 MOUSE	Q8VCA5 mus musculus
6	676.5	24.9	437	1 TMS4 HUMAN	Q9NR84 homo sapien
7	665.5	24.5	455	1 TMS5 MOUSE	Q98072 bos taurus
8	660	24.3	1035	1 TMS5 BOVIN	P98074 mus musculus
9	651.5	24.0	457	1 TMS5 HUMAN	Q9H383 homo sapien
10	647	23.8	1019	1 ENTX HUMAN	P98073 homo sapien
11	638	23.5	1069	1 ENTX MOUSE	P97435 mus musculus
12	634	23.3	1034	1 ENTX PIG	P98074 sus scrofa
13	601	22.1	1042	1 COR1 HUMAN	Q9Y5G5 homo sapien
14	577.5	21.3	1113	1 COR1 MOUSE	Q92319 mus musculus
15	569.5	21.0	638	1 KAL HUMAN	Q92319 mus musculus
16	568.5	20.9	855	1 TMS6 MOUSE	P56677 mus musculus
17	565.5	20.8	417	1 HEP5 HUMAN	P05981 homo sapien
18	558.5	20.6	855	1 ST14 HUMAN	Q9Y5Y6 homo sapien
19	556	20.5	638	1 KAL RAT	P14272 rattus norv
20	556	20.5	611	1 TMS6 MOUSE	Q9CB10 mus musculus
21	547	20.1	436	1 HEP5 MOUSE	Q35453 mus musculus
22	538	19.8	638	1 KAL MOUSE	P26262 mus musculus
23	536.5	19.7	611	1 TMS6 HUMAN	Q81U80 homo sapien
24	533	19.6	625	1 FALL HUMAN	P03951 homo sapien
25	529.5	19.5	416	1 HEP5 RAT	Q05511 rattus norv
26	527.5	19.4	418	1 HATT HUMAN	O60235 homo sapien
27	517.5	19.0	612	1 PLMN MOUSE	P0918 mus musculus
28	504.5	18.6	324	1 TEST MOUSE	Q9JH77 mus musculus
29	501	18.4	421	1 ACRO HUMAN	P10323 homo sapien
30	496	18.3	422	1 DES1 HUMAN	Q9UL52 homo sapien
31	490.5	18.1	415	1 ACRO PIG	P98001 sus scrofa
32	489.5	18.0	436	1 ACRO MOUSE	P23578 mus musculus
33	486.5	17.9	761	1 NEUR MOUSE	O08762 mus musculus

ALIGNMENTS

RESULT 1

TMS2_HUMAN	TMS2_HUMAN	STANDARD;	PRT;	492 AA.
ID	O15393; Q9BXX1;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Transmembrane protease, serine 2 precursor (EC 3.4.21.-).			
GN	TPRSS2 OR PRS10.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97468144; PubMed=9325052;			
RA	Paoloni-Giacobino A., Chen H., Peitsch M.C., Rossier C.,			
RA	Antonarakis S.E.;			
RT	"Cloning of the TPRS2 gene, which encodes a novel serine protease			
RT	with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3.";			
RL	Genomics 44:309-320(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21309069; PubMed=11414763;			
RA	Teng D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;			
RT	"Mutation analyses of 288 candidate genes in human tumor cell lines.";			
RL	Genomics 74:352-364(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A., AND MUTAGENESIS.			
RX	MEDLINE=21139112; PubMed=11245484;			
RA	Afar D.E.H., Vivanco I., Rubert R.S., Kuo J., Chen E., Saffran D.C.,			
RA	Raitano A.B., Jakobovits A.;			
RT	"Catalytic cleavage of the androgen-regulated TPRS2 protease results			
RT	in its secretion by prostate and prostate cancer epithelia.";			
RL	Cancer Res. 61:1686-1692(2001).			
RN	[4]			
RP	TISSUE SPECIFICITY.			
RX	MEDLINE=21104370; PubMed=11169526;			
RA	Vaaxala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;			
RT	"Expression of transmembrane serine protease TPRS2 in mouse and			
RT	human tissues.";			
RL	J. Pathol. 193:134-140(2001).			
CC	-I- SUBCELLULAR LOCATION: Type II membrane protein. Activated by			
CC	cleavage and secreted.			
CC	-I- TISSUE SPECIFICITY: Expressed strongly in small intestine. Also			
CC	expressed in prostate, colon, stomach and salivary gland.			
CC	-I- SIMILARITY: Belongs to peptidase family S1.			
CC	-I- SIMILARITY: Contains 1 LDL-receptor class A domain.			
CC	-I- SIMILARITY: Contains 1 SRCR domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/)			

34	477	17.6	790	1 PLMN PIG	P06867 sus scrofa
35	475.5	17.5	431	1 ACRO_RABIT	P45038 coryctolagus
36	472.5	17.4	810	1 PLMN_MACMU	P12545 macaca mulia
37	471	17.3	290	1 PR27 HUMAN	Q9BQR3 homo sapien
38	470.5	17.3	437	1 ACRO RAT	P29293 rattus norv
39	470.5	17.3	810	1 PLMN_ERIEU	Q29485 erinaceus e
40	469.5	17.3	311	1 TRYX_MOUSE	Q9QU17 mus musculus
41	467.5	17.2	321	1 TRYX_HUMAN	Q9RR22 homo sapien
42	463	17.0	273	1 TRYX_SHEEP	Q9X8M2 ovis aries
43	462	17.0	317	1 BSS4_HUMAN	Q93ZM4 homo sapien
44	461	17.0	343	1 PSS6_HUMAN	Q16651 homo sapien
45	460.5	16.9	306	1 BSS4_MOUSE	Q9ER10 mus musculus

```

or send an email to license@sib-sib.ch).
-----
CC  EMBL; U75329; AAC51784.1; -
DR  EMBL; AF123453; AAD37117.1; -
DR  EMBL; AF270487; AAK29280.1; -
DR  HSP; P00763; IDPO.
DR  MEROPS; S01.247; -.
DR  Genew; HGNC:11876; TMPRSS2.
DR  MIM; 602060; -.
DR  GO; GO:0005887; C: integral to plasma membrane; TAS.
DR  GO; GO:0008236; F: serine-type peptidase activity; TAS.
DR  InterPro; IPR009003; Cys Ser. trypsin.
DR  InterPro; IPR002172; LDL_receptor_A.
DR  InterPro; IPR001254; Peptidase_S1.
DR  InterPro; IPR001314; Peptidase_S1A.
DR  InterPro; IPR001190; Srcr_receptor.
DR  Pfam; PF00089; trypsin; 1.
DR  PRINTS; PR00722; CHYMOTRYPSIN.
DR  SMART; SM00192; LDLA; 1.
DR  SMART; SM00202; SR; 1.
DR  SMART; SM00200; TRYD_SPC; 1.
DR  PROSITE; PS01209; LDLA_1; 1.
DR  PROSITE; PS00888; LDLA_2; 1.
DR  PROSITE; PS00420; SRCR_1; FALSE NEG.
DR  PROSITE; PS0287; SRCR_2; 1.
DR  PROSITE; PS0240; TRYPSIN_DOM; 1.
DR  PROSITE; PS00134; TRYPSIN_HIS; 1.
DR  PROSITE; PS00135; TRYPSIN_SER; 1.
KW  Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
KW  Polymorphism.
FT  CHAIN 1 255 TRANSMEMBRANE PROTEASE, SERINE 2, NON-
FT  CHAIN 256 492 CATALYTIC CHAIN.
FT  CHAIN 492 TRANSMEMBRANE PROTEASE, SERINE 2,
FT  CHAIN CATALYTIC CHAIN.
FT  DOMAIN 1 84 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT  DOMAIN 85 105 (POTENTIAL).
FT  DOMAIN 106 492 EXTRACELLULAR (POTENTIAL).
FT  DOMAIN 112 149 LDL-RECEPTOR CLASS A.
FT  DOMAIN 150 242 SRCR.
FT  DOMAIN 256 492 SERINE PROTEASE.
FT  ACT_SITE 296 296 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT  ACT_SITE 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT  ACT_SITE 441 441 CHARGE RELAY SYSTEM.
FT  SITE 255 256 CLEAVAGE (POTENTIAL).
FT  DISULFID 113 126 BY SIMILARITY.
FT  DISULFID 120 139 BY SIMILARITY.
FT  DISULFID 133 148 BY SIMILARITY.
FT  DISULFID 172 231 BY SIMILARITY.
FT  DISULFID 185 241 BY SIMILARITY.
FT  DISULFID 244 365 INTERCHAIN (BY SIMILARITY).
FT  DISULFID 281 297 BY SIMILARITY.
FT  DISULFID 410 426 BY SIMILARITY.
FT  DISULFID 437 455 BY SIMILARITY.
FT  CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  VARIANT 449 449 K -> N (in dbSNP:1056602).
FT  MUTAGEN 255 255 R->Q: LOSS OF CLEAVAGE.
FT  MUTAGEN 441 441 S->A: LOSS OF ACTIVITY.
FT  CONFLICT 160 160 M -> V (IN REF. 3).
FT  CONFLICT 242 242 I -> L (IN REF. 1).
FT  CONFLICT 329 329 E -> Q (IN REF. 1).
FT  CONFLICT 489 491 RAD -> KAN (IN REF. 1).
FT  SEQUENCE 492 AA; 53891 MW; CAB44FD174A9076B CRC64;
Query Match 99.9%; Score 2714; DB 1; Length 492;
Best Local Similarity 99.8%; Pred. No. 2,6e-193;
Matches 491; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MALNCGSPPAIGFYENHGQYQENYPACPTVPTVEVHPAQYYPFQYAPRVLTQA 60
DB 1 MALNCGSPPAIGFYENHGQYQENYPACPTVPTVEVHPAQYYPFQYAPRVLTQA 60

```

```

QY 61 SNPVVCTQKPSGTVCTSKTKKALCITLTLGFLVGAALAAAGLLKFMGSKCSNSGIEC 120
DB 61 SNPVVCTQKPSGTVCTSKTKKALCITLTLGFLVGAALAAAGLLKFMGSKCSNSGIEC 120
QY 121 DSSGTCINPSNWCDSVSHCPGDEENRCVRLYGNFLOVYSSQKSKWHHPVCQDDNNENY 180
DB 121 DSSGTCINPSNWCDSVSHCPGDEENRCVRLYGNFLOVYSSQKSKWHHPVCQDDNNENY 180
QY 181 GRAACRDMGYKXNFYSSOGIVDDSGSTSFMKLNTSAGNVDIYKKLHYSDACSKKAVVSLR 240
DB 181 GRAACRDMGYKXNFYSSOGIVDDSGSTSFMKLNTSAGNVDIYKKLHYSDACSKKAVVSLR 240
QY 241 CIACGVNLNSRQRIYVGSALPGAWPQVSLHVQNVHVCVGGSIITPEWIVTAACHVEK 300
DB 241 CIACGVNLNSRQRIYVGSALPGAWPQVSLHVQNVHVCVGGSIITPEWIVTAACHVEK 300
QY 301 PLNNPWHMTAFAGILRQSFMYGAGYQVKEVIFHPNYDSKTKNDIALMKLQKPLFNDL 360
DB 301 PLNNPWHMTAFAGILRQSFMYGAGYQVKEVIFHPNYDSKTKNDIALMKLQKPLFNDL 360
QY 361 VKPVCLPNPGMQLPEQLCWISGWGATEBKTKSEVLNAKVLLETORCNSRYVDNLI 420
DB 361 VKPVCLPNPGMQLPEQLCWISGWGATEBKTKSEVLNAKVLLETORCNSRYVDNLI 420
QY 421 TPAMICAGFLQGNVDSQCGDSGGLVTSKNNIWLIGTSGWGGCAKAYRPGVYGNVMVF 480
DB 421 TPAMICAGFLQGNVDSQCGDSGGLVTSKNNIWLIGTSGWGGCAKAYRPGVYGNVMVF 480
QY 481 TDWYRQWRADG 492
DB 481 TDWYRQWRADG 492
RESULT 2
TMS2_MOUSE
ID TMS2_MOUSE STANDARD; PRT; 490 AA.
AC Q9JQ8; Q9JTK4; Q9QY82;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transmembrane protease, serine 2 (EC 3.4.21.-) (Epitheliasin) (Plasmic
DE transmembrane protein X).
GN TMPRSS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=21104370; PubMed=11169526;
RX STRAIN=BALB/c;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Viikko P.T.;
RT "Expression of transmembrane serine protease TMPRSS2 in mouse and
RT human tissues.";
RL J. Pathol. 193:134-140 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Han J., Kim S.;
RT "Putative transmembrane protease X.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=BALB/c;
RX MEDLINE=20148617; PubMed=10683448;
RA Jacquinet E.J., Rao N.V., Rao G.V., Hoidal J.R.;
RT "Cloning, genomic organization, chromosomal assignment and expression
RT of a novel mosaic serine proteinase: epitheliasin.";
RL FEBS Lett. 468:93-100 (2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=Mammary gland;
RX MEDLINE=22386257; PubMed=12477932;

```

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore I., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzay D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Activated by
CC cleavage and secreted (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed mainly in prostate and kidney.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; AF199362; AAF97867.1; -;
DR ENBL; AF243500; AAF64186.1; -;
DR ENBL; AF113596; AAF21308.1; -;
DR ENBL; BC038393; AAR38393.1; -;
DR HSP; P00763; LDPO.
DR MEROPS; S01.247; -;
DR MGD; MGI:1354381; Tmorss2.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00530; SRCR; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; Ldla; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00202; Tryp_Spc; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS0068; LDLRA_2; 1.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen.
FT CHAIN 1 253
FT TRANSMEMBRANE PROTEASE, SERINE 2, NON-
FT CATALYTIC CHAIN.
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT CYTOPLASMIC (POTENTIAL).
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT EXTRACELLULAR (POTENTIAL).
FT LDL-RECEPTOR CLASS A.
FT SRCR.
FT SERINE PROTEASE.
FT 254 490

FT ACT_SITE 294 343 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 343 343 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 439 439 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 253 125 CLEAVAGE (POTENTIAL).
FT DISULFID 76 125 BY SIMILARITY.
FT DISULFID 119 138 BY SIMILARITY.
FT DISULFID 132 147 BY SIMILARITY.
FT DISULFID 171 230 BY SIMILARITY.
FT DISULFID 184 240 BY SIMILARITY.
FT DISULFID 243 363 INTERCHAIN (BY SIMILARITY).
FT DISULFID 279 295 BY SIMILARITY.
FT DISULFID 408 434 BY SIMILARITY.
FT DISULFID 435 463 BY SIMILARITY.
FT CARBOHYD 111 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 122 122 S -> L (IN REF. 3).
FT CONFLICT 178 178 S -> N (IN REF. 3).
FT CONFLICT 320 320 Y -> H (IN REF. 1).
FT CONFLICT 474 474 N -> D (IN REF. 1).
SQ SEQUENCE 490 AA; 53479 MW; 07D2B03EA4D8A1A9 CRC64;
Query Match 77.9%; Score 2117; DB 1; Length 490;
Best Local Similarity 78.4%; Pred. No. 3.4e-149;
Matches 395; Conservative 41; Mismatches 63; Indels 2; Gaps 2;
QY 1 MALNGSPPAIGPYENHYQENPYPAQPTVPTVEVHPAQQYPSVPCYARVLTOA 60
DB 1 MALNGSPGIGPCYENHYQSEHICPPRPVPAENGYNLYPAQYPSVPQYAPRITTOA 60
QY 61 SFPVCTCPKPSGTVCTSKTKKALCITLTGLTFLVGAALAGLWKFMGKSCNSGIEC 120
DB 61 STSVITHPKS-SGAPCTSKSKSLCALALGTLTGAAVAALLWRFWDNSCSTSEMEC 119
QY 121 DSGGTCINPNCWDCGSHCPGEGDENRCVLYGPNFILOYVSSORKSHWPCQDDWNEY 180
DB 120 GSGGTCISLWCDGVAHCPNGEDENRCVLYGSGFILOYVSSORKAWYPVCCDDWSEY 179
QY 181 GRAACRDGYNKFNFYSSQIGIVDDSGSTSMKLNATAGNVDYKLYHSDACSSKAVSLR 240
DB 180 GRAACKDGYKNFYSSQIGIPDQSGATSMKLNSSGNDVLYKLYHSDSCSSRMVSLR 239
QY 241 CTACGNVNSSRQSRIVGEGSALPGAMPQVSLHVNQVHVGCGSIITPEWITVAHCVK 300
DB 240 CTECGVR-SVKGQSRIVGGLNASPGDWPQVSLHVQGVHVGCGSIITPEWITVAHCV 298
QY 301 PLNNPWHWTAFAGILRQSFYMGYQYQYKVISHPNVDSTKKNNDIALMKLQPLTFNDL 360
DB 299 PLSGRYWTAFAGILRQSLMFYSGRHOVEKVISHPNVDSTKKNNDIALMKLQPLTFNDL 358
QY 361 VPKVCLPNPQMMLOPEQLCWISGWGATEKGTSEVLNAKVLIETQRCNSRVYDNL 420
DB 359 VPKVCLPNPQMMLOPEQLCWISGWGATEKGTSEVLNAKVLIETQRCNSRVYDNL 418
QY 421 TPAMICAGFLQGVDSQCGSGPLVTSKNINWLGITSGWGCACAYRPGVYGVNVVF 480
DB 419 TPAMICAGFLQGVDSQCGSGPLVTSKNINWLGITSGWGCACAYRPGVYGVNVVF 478
QY 481 TDWITVROMRAD 491
DB 479 TDWITVROMRAD 489
RESULT 3
TMS3_MOUSE
ID_TMS3_MOUSE STANDARD; PRT; 453 AA.
AC Q8K1T0; Q8VDE0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transmembrane protease, serine 3 (EC 3.4.21.-).
GN TMRPS3.
OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND FUNCTION IN ENAC
 RP CLEAVAGE.
 RX MEDLINE=22281255; PubMed=12393794;
 RA Guilponi M., Vuagniaux G., Wattenhofer M., Shibuya K., Vazquez M.,
 RA Douchery L., Scamuffa N., Guida E., Okui M., Rossier C., Hancock M.,
 RA Bucher K., Raymond A., Hummel E., Marzella P.L., Kudoh J.,
 RA Shimizu N., Scott H.S., Antonarakis S.E., Rossier B.C.;
 RT "The transmembrane serine protease (TMPRSS3) mutated in deafness
 RL DNFB8/10 activates the epithelial sodium channel (ENaC) in vitro.";
 RL Hum. Mol. Genet. 11:2829-2836(2002).
 CC -!- FUNCTION: Probable protease. Seems to be capable of activating
 CC ENaC.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
 CC reticulum.
 CC -!- TISSUE SPECIFICITY: Expressed in the spiral ganglion, the cells
 CC supporting the organ of Corti and the stria vascularis.
 CC -!- PTM: Undergoes autolytic activation.
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 CC -!- SIMILARITY: Contains 1 LDL-receptor class A domain.
 CC -!- SIMILARITY: Contains 1 SRCR domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC -----
 CC EMBL: AJ429216; CAD22137.1; --
 CC EMBL: AJ300738; CAC83350.1; --
 CC HSSP: P00761; IANI.
 CC MGD: MGI:2155445; Tmprs3.
 CC InterPro: IPR009003; Cys_Ser_trypsin.
 CC InterPro: IPR002172; LDL_receptor_A.
 CC InterPro: IPR001254; Peptidase_S1.
 CC InterPro: IPR001314; Peptidase_S1A.
 CC InterPro: IPR001130; Srcr_receptor.
 CC Pfam: PF00057; ldl_recept_a; 1.
 CC Pfam: PF00089; trypsin; 1.
 CC PRINTS: PR00722; CHYMOTRYPSIN.
 CC SMART: SM00192; LDLa; 1.
 CC SMART: SM00202; SR; 1.
 CC SMART: P00020; Tryp_SPC; 1.
 CC PROSITE: PS01209; LDLRA_1; 1.
 CC PROSITE: PS00668; LDLRA_2; 1.
 CC PROSITE: PS0287; SRCR_2; 1.
 CC PROSITE: PS0240; TRYPsin_DOM; 1.
 CC PROSITE: PS00134; TRYPsin_HIS; 1.
 CC PROSITE: PS00135; TRYPsin_SER; 1.
 CC Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
 KW Endoplasmic reticulum.
 FT DOMAIN 1 48
 FT TRANSMEM 49 69
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT EXTRACELLULAR (POTENTIAL).
 FT LDL-RECEPTOR CLASS A.
 FT SRCR.
 FT SERINE PROTEASE.
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 257 257
 FT ACT_SITE 304 304
 FT ACT_SITE 400 400
 FT SITE 215 217
 FT CLEAVAGE (POTENTIAL).
 FT BY SIMILARITY.
 FT DISULFID 73 85
 FT DISULFID 79 98
 FT BY SIMILARITY.
 FT DISULFID 92 107
 FT BY SIMILARITY.
 FT DISULFID 129 194
 FT BY SIMILARITY.
 FT DISULFID 142 204
 FT BY SIMILARITY.
 FT DISULFID 207 324
 FT BY SIMILARITY.

FT DISULFID 242 258 BY SIMILARITY.
 FT DISULFID 338 406 BY SIMILARITY.
 FT DISULFID 369 385 BY SIMILARITY.
 FT DISULFID 396 424 BY SIMILARITY.
 FT CARBOHYD 221 221 N-LINKED (GLCNAC... (POTENTIAL).
 FT CONFLICT 117 117 L -> H (IN REF. 1; CAC83350).
 FT CONFLICT 246 246 V -> I (IN REF. 1; CAC83350).
 SQ SEQUENCE 453 AA; 49491 MW; 1ABCBF10AF6E1EF6 CRC64;
 Query Match 32.4%; Score 880; DB 1; Length 453;
 Best Local Similarity 42.9%; Pred. No. 8.5e-58;
 Matches 187; Conservative 64; Mismatches 159; Indels 26; Gaps 12;
 QY 69 PKPSGTVCTSKTKKALCI-----TLTLGLTFLVGAALAGLLMKFMGSKNSGIEDSSG 124
 DB 28 FVAPDGDVAQAQILSLPLKFFPIIVIGIILALALALGGLGIFH---DCSGK-YRCHSSF 83
 QY 125 TCINPSNWCDCVSHCPGGEDENCVRLYGNFVLQVYSSQKSHHPVCCDDNNENYGRAA 184
 DB 84 KCIELTARCDGVSDCKNAEDYRCVSGORALQVFTA--AAWRTMCSDDMKSHYAKIA 441
 QY 185 CRDMGYKNFYSSQGI---VDPSSGTSFMKLNTSAGNVDIYKLYHS----DACSSKAV 236
 DB 142 CAQLGFP-SYVSDHLRVDALEREQFGDFVSIINELLSD-DKVTALHHSVYMRGCTSGHV 199
 QY 237 VSLRCIACGVNLSSRSRIRVGGESALPGAPWPOVSHVQNVHVCGSIIITPEWIVTAAK 296
 DB 200 VILKCSACGRTGYS--PRIVGNNMSLTQWPOVSLQFGYHLCGGSVITPLWITAAH 257
 QY 297 CVEKPLNPHWHTAFAGILRQSFMYGA-GYQVEKVISHPNYDSKTKNNDIAMKLQKPL 355
 DB 258 CV-YDLVHPKSWTVQVGLV--SLMDSVPVSHLVEKIYHSHKPKPLGNDIALMKLSEPL 314
 QY 356 TFNDLVKPVCLPNPQMLQPEOLCISGWCATBEKGTSEVLNAAKVLLETQRCNSRYV 415
 DB 315 TFDTIQICLIPNSEENFPDGKLCWISGWATDGGDAPVLNAAVPLISNKICNRDV 374
 QY 416 YDNLTITPAMTCAGFLQNVDSQCGDGLVTSKNNIMWLIGTSGWGGCAKAYRPGVYG 475
 DB 375 YGGIISPSMLCAGYKGGVDSQCGDGLVTCQERBLMKLVGATSGFGICAEVKNKPGVYT 434
 QY 476 NMVUFTDNIYQMRAD 491
 DB 435 RITSFLDWIHEQLERD 450
 RESULT 4
 TMS3 HUMAN
 ID TMS3 HUMAN STANDARD; PRT; 454 AA.
 AC P57727.
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Transmembrane protease, serine 3 (EC 3.4.21.-) (Serine protease
 DE TADG-12) (Tumor associated differentially-expressed gene-12 protein).
 GN TMPRSS3 OR TADG12 OR ECHOS1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS A AND T).
 RC TISSUE=Ovarian carcinoma;
 RX MEDLINE=20521358; PubMed=11068177;
 RA Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N.,
 RA Wang Y., Farnley T.H., O'Brien T.J.;
 RT "Ovarian tumor cells express a novel multi-domain cell surface serine
 RT protease".
 RL Biochim. Biophys. Acta 1502:337-350(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS A; B AND D), AND VARIANT IL5-53.
 RX MEDLINE=20578749; PubMed=1137999;
 RX Scott H.S., Kudoh J., Wattenhofer M., Shibuya K., Berry A., Chraist R.,

RA	Guipponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S.,	CC	-1- SIMILARITY: Contains 1 SRCR domain.
RA	Younus F., Mehdi S.O., Radhakrishna U., Papasavvas M.P., Gehrig C.,	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
RA	Rossier C., Korostishevsky M., Gal A., Shimizu N., Bonne-Tamir B.,	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
RA	Antonarakis S.E.,	CC	the European Bioinformatics Institute. There are no restrictions on its
RT	"Insertion of beta-satellite repeats identifies a transmembrane	CC	use by non-profit institutions as long as its content is in no way
RT	protease causing both congenital and childhood onset autosomal	CC	modified and this statement is not removed. Usage by and for commercial
RT	recessive deafness.";	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
RL	Nat. Genet. 27:59-63(2001).	CC	or send an email to license@isb-sib.ch).
RL	[3]	CC	
RP	SUBCELLULAR LOCATION, AND FUNCTION IN ENAC CLEAVAGE.	CC	
RP	MEDLINE=22281258; PubMed=12333794;	DR	EMBL; AF201380; BAB20077.1; -
RP		DR	EMBL; AB038157; BAB20077.1; -
RA	Guipponi M., Vugniatoux G., Wattenhofer M., Shibuya K., Vazquez M.,	DR	EMBL; AB038158; BAB20078.1; -
RA	Dougherty L., Scamuffa N., Guida E., Okui M., Rossier C., Hancock M.,	DR	EMBL; AB038159; BAB20079.1; -
RA	Buchet K., Raymond A., Hummler B., Marzella P.L., Kudoh J.,	DR	EMBL; AB038160; BAB20080.1; -
RA	Shimizu N., Scott H.S., Antonarakis S.E., Rossier C.,	DR	HSP; P00763; IDPO.
RA	"The transmembrane serine protease (TMPRSS3) mutated in deafness	DR	MEROPS; S01.075; -
RT	DFNB8/10 activates the epithelial sodium channel (ENaC) in vitro.";	DR	Genew; HGNC:11877; TMPRSS3.
RL	Hum. Mol. Genet. 11:2829-2836(2002).	DR	MIM; 605511; -
RL	[4]	DR	MIM; 601072; -
RP	VARIANTS DFNB8/DFNB10 CYS-251 AND LEU-404.	DR	MIM; 605316; -
RP	MEDLINE=21334482; PubMed=11462234;	DR	GO; GO:0016021; C:integral to membrane; NAS.
RP		DR	GO; GO:0004252; F:serine-type endopeptidase activity; NAS.
RA	Masmoudi S., Antonarakis S.E., Schwede T., Ghorbel A.M., Gratri M.,	DR	GO; GO:000508; P:proteolysis and peptidolysis; NAS.
RA	Pappasavvas M.P., Dirra M., Elgail-Boulila A., Wattenhofer M.,	DR	InterPro; IPR009003; Cys_Ser_trypsin.
RA	Rossier C., Scott H.S., Avadi H., Guipponi M.,	DR	InterPro; IPR001172; LDL_receptor A.
RT	"Novel missense mutations of TMPRSS3 in two consanguineous Tunisian	DR	InterPro; IPR001254; Peptidase S1A.
RT	families with non-syndromic autosomal recessive deafness.";	DR	InterPro; IPR001190; Srrc receptor.
RL	Hum. Mutat. 18:101-108(2001).	DR	InterPro; IPR001314; Peptidase S1A.
RL	[5]	DR	InterPro; IPR00057; ldl_recept_a; 1.
RP	VARIANTS DFNB8/DFNB10 TRP-109; PHE-194 AND ARG-407, AND VARIANTS	DR	Pfam; PF000089; trypsin; 1.
RP	ILE-53; SER-111 AND VAL-253.	DR	PRINTS; PR00722; CHYMOTRYPSIN.
RP	MEDLINE=21317610; PubMed=11424922;	DR	SMART; SM00192; LDLA; 1.
RA	Ben-Yosef T., Wattenhofer M., Riazuddin S., Ahmed Z.M., Scott H.S.,	DR	SMART; SM00202; SR; 1.
RA	Kudoh J., Shibuya K., Antonarakis S.E., Bonne-Tamir B.,	DR	PROSITE; PS01209; LDLRA_1; 1.
RA	Radhakrishna U., Naz S., Ahmed Z., Riazuddin S., Pandya A.,	DR	PROSITE; PS00668; LDLRA_2; 1.
RA	Nance W.E., Wilcox E.R., Friedman T.B., Morell R.J.,	DR	PROSITE; PS00420; SRCR_1; FALSE_NEG.
RA	"Novel mutations of TMPRSS3 in four DFNB8/B10 families segregating	DR	PROSITE; PS0287; SRCR_2; 1.
RT	congenital autosomal recessive deafness.";	DR	PROSITE; PS0240; TRYPSIN_DOM; 1.
RL	J. Med. Genet. 38:396-400(2001).	DR	PROSITE; PS00134; TRYPSIN_HIS; 1.
RL	[6]	DR	PROSITE; PS00135; TRYPSIN_SER; 1.
RP	VARIANT DFNB8/DFNB10 GLY-103, AND VARIANTS ASN-173 AND THR-426.	KW	Hydrolase; serine protease; Transmembrane; Signal-anchor; Zymogen;
RP	MEDLINE=21904597; PubMed=11907649;	KW	Endoplasmic reticulum; Deafness; Alternative splicing;
RA	Wattenhofer M., Di Iorio V., Rabinet R., Dougherty L., Pampas A.,	FT	Disease mutation; Polymorphism.
RA	Schwede T., Montserrat-Sentis B., Arbones L., Iliades T.,	FT	DOMAIN 1 48
RA	Pasquadrile A., D'Amelio M., Alwan S., Rossier C., Dahl H.-H.M.,	FT	TRANSMEM 49 69
RA	Petersen M.B., Estrivall X., Gasparini P., Scott H.S.,	FT	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
RA	Antonarakis S.E.,	FT	(POTENTIAL).
RT	"Mutations in the TMPRSS3 gene are a rare cause of childhood	FT	EXTRACELLULAR (POTENTIAL).
RT	non-syndromic deafness in Caucasian patients.";	FT	LDL-RECEPTOR CLASS A.
RL	J. Mol. Med. 80:124-131(2002).	FT	SRCR.
RL	[7]	FT	SERINE PROTEASE.
CC	-1- FUNCTION: Probable protease. Seems to be capable of activating	FT	CHARGE RELAY SYSTEM (BY SIMILARITY).
CC	ENAC.	FT	CHARGE RELAY SYSTEM (BY SIMILARITY).
CC	-1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic	FT	CHARGE RELAY SYSTEM (BY SIMILARITY).
CC	reticulum.	FT	CLEAVAGE (POTENTIAL).
CC	-1- ALTERNATIVE PRODUCTS:	FT	BY SIMILARITY.
CC	Event=Alternative splicing; Named isoforms=4;	FT	BY SIMILARITY.
CC	Name=A;	FT	BY SIMILARITY.
CC	IsoId=P57727-1; Sequence=Displayed;	FT	BY SIMILARITY.
CC	Name=B; Synonyms=C;	FT	BY SIMILARITY.
CC	IsoId=P57727-2; Sequence=VSP_005391;	FT	BY SIMILARITY.
CC	Name=D;	FT	BY SIMILARITY.
CC	IsoId=P57727-3; Sequence=VSP_005392;	FT	BY SIMILARITY.
CC	Name=T; Synonyms=Truncated, TAGD-12V;	FT	BY SIMILARITY.
CC	IsoId=P57727-4; Sequence=VSP_005393, VSP_005394;	FT	BY SIMILARITY.
CC	-1- TISSUE SPECIFICITY: Expressed in many tissues including fetal	FT	BY SIMILARITY.
CC	cochlea. Isoform T is found at increased levels in some	FT	BY SIMILARITY.
CC	carcinomas.	FT	BY SIMILARITY.
CC	-1- PTM: Undergoes autolytic activation.	FT	BY SIMILARITY.
CC	-1- DISEASE: Defects in TMPRSS3 are a cause of childhood-onset	FT	N-LINKED (GLCNAC..) (POTENTIAL).
CC	autosomal recessive neurosensory deafness 8 (DFNB8) [MIM:601072].	FT	Missing (in isoform B).
CC	-1- DISEASE: Defects in TMPRSS3 are a cause of congenital autosomal	FT	/FTId=VSP_005391.
CC	recessive neurosensory deafness 10 (DFNB10) [MIM:605316].	FT	EMIQVCLPNSSENFDPKVCWTSWGMGATEDGADSPVLN
CC	-1- SIMILARITY: Belongs to peptidase family S1.	FT	HAAPVLSNKTCHNRDVGIIISPLMCAVLTGVDSCOG
CC	-1- SIMILARITY: Contains 1 LDL-receptor class A domain.	FT	DSCGPLVCQBRRLKLVGATSGICAVNKPQVYTVTSF
CC		FT	LDWTHIQNERDKT -> GISGSLGCSAALPLFQEDLQLLI

DR Pfam; PF00057; ldl_recept a; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDRA_1; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00202; SR; 1.
 DR PROSITE; PS01209; LDRA_1; FALSE NEG.
 DR PROSITE; PS00668; LDRA_2; FALSE NEG.
 DR PROSITE; PS00420; SRCR_1; FALSE NEG.
 DR PROSITE; PS00287; SRCR_2; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Transmembrane; Signal-anchor.
 FT DOMAIN 1 32 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 33 53 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 54 437 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 61 93 LDL-RECEPTOR CLASS A.
 FT DOMAIN 94 204 SRCR.
 FT DOMAIN 205 437 SERINE PROTEASE.
 FT ACT_SITE 245 245 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 290 290 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 387 387 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 204 205 CLEAVAGE (POTENTIAL).
 FT DISULFID 64 83 BY SIMILARITY.
 FT DISULFID 77 92 BY SIMILARITY.
 FT DISULFID 127 183 BY SIMILARITY.
 FT DISULFID 140 193 BY SIMILARITY.
 FT DISULFID 196 310 BY SIMILARITY.
 FT DISULFID 230 246 BY SIMILARITY.
 FT DISULFID 356 372 BY SIMILARITY.
 FT DISULFID 383 410 BY SIMILARITY.
 FT CARBOHYD 130 178 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 1 31 MLDPSDQPSLNSLDVLPKRPDPMETFRK -> MSNPCA
 FT NPYSPWRSES (IN REF. 2).
 SQ SEQUENCE 437 AA; 48204 MW; 351B2FD4A8657B12 CRC64;
 Query Match 24.9%; Score 676.5; DB 1; Length 437;
 Best Local Similarity 39.1%; Pred. No. 96-43;
 Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;
 QY 133 CDGVSHCPGGEENRCVRLY--GP-----NFILQVSSQKSHHPVQCDQDNNENYG 181
 DB 77 CDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLSATGNWFSACFDNFTALA 136
 QY 182 RAACRDNGY--KNNFYSSQ-----GIVDPSGTSFMKLNTSAGNVDIYKLYHSDACS 232
 DB 137 ETACROMGYSSKPTFAVEIGPDQLDVVEITENSQELRMNNSG-----PCL 184
 QY 233 SKAVVSLRCIACGVNLNSRQRIYVGGESALPGAPWQVSLHVQNVHVCVGGSIITPEWIV 292
 DB 185 SGLSVSLHCLACGKSL---KTPRVVGGEEASVDSWPQVSIQYDKQKRVCGSILDPHWVL 241
 QY 293 TAAHCEKFLNPNHWTAFAGILR-QSEMFYAGYQVEKVI---SHPNYDSTKNNNDIAL 348
 DB 242 TAAHCFKH-TDVFNNKVRAGSDKLGSP---PSLAVAKIIIEFPNMY---PKNDIAL 293
 QY 349 MKLQKPLTNDLVKPVCLNPGMMLQPEQLCHISGATEEK-GKTSEVLNAKVLLET 407
 DB 294 MKLQPLTTEGTVRPICLFFPEELTPATPLWICGFTKQNGKMSDILLQASVQVIDS 353
 QY 408 QRCNRYVYDNLITPAMICAGFLQGNVDSGQSGGLVTSKNNIWMILGDTWSGCGAK 467
 DB 354 TRCNADDAQGEVTEKQWMCAGIPFEGVDTCQDGGGLMYQSDQ-WHVVGIVSNVGYCGG 412
 QY 468 AVRFVYGVNMYFTWIKPQWAD 491
 DB 413 PSTPGVYTKVSAYLNNWYNWKA 436

RESULT 7

TM55 MOUSE
 ID Q9ER04; Q9ER02; Q9ER03; PRT; 455 AA.
 AC Q9ER04; Q9ER02; Q9ER03; PRT; 455 AA.
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2003 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
 GN TMPSRS.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]_TaxID=10090;
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RC TISSUE=Brain;
 RA Mitsui S., Yamaguchi N.;
 RT "cDNA cloning of mouse spinesin";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RC TISSUE=Brain;
 RA Mitsui S., Yamaguchi N.;
 RT "Molecular cloning of mouse type 4 spinesin";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC 1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
 CC 1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Name=4;
 CC IsoId=Q9ER04-1; Sequence=Displayed;
 CC Name=1;
 CC IsoId=Q9ER04-2; Sequence=VSP_005397, VSP_005398;
 CC Name=2;
 CC IsoId=Q9ER04-3; Sequence=VSP_005395;
 CC Name=3;
 CC IsoId=Q9ER04-4; Sequence=VSP_005396;
 CC 1- SIMILARITY: Belongs to peptidase family S1.
 CC 1- SIMILARITY: Contains 1 SRCR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC
 DR EMBL; AB016229; BAB20276.1; -
 DR EMBL; AB016230; BAB20277.1; -
 DR EMBL; AB016423; BAB20278.1; -
 DR EMBL; AB041037; BAB40328.1; -
 DR HSSP; P00763; 1DPO.
 DR MEROPS; S01.313; -
 DR MGD; MGI:1933407; Tmpres5.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srcr_receptor.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; TRYP_SPC; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE; PS0287; SRCR_2; 1.
 KW Hydrolase; Serine protease; Transmembrane; Signal-anchor;
 KW Glycoprotein; Alternative splicing.
 FT DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 71 455 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 112 207 SRCR.
 FT DOMAIN 218 455 SERINE PROTEASE.

```

FT ACT_SITE 258 258 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 217 218 CLEAVAGE (POTENTIAL).
FT DISULFID 135 196 BY SIMILARITY.
FT DISULFID 148 206 BY SIMILARITY.
FT DISULFID 209 328 BY SIMILARITY.
FT DISULFID 243 259 BY SIMILARITY.
FT DISULFID 374 390 BY SIMILARITY.
FT DISULFID 401 429 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 144 Missing (in isoform 2).
FT VARSPLIC 1 10 Missing (in isoform 3).
FT VARSPLIC 1 182 Missing (in isoform 1).
FT VARSPLIC 183 192 /FTid=VSP 005397.
FT VARSPLIC 325 325 /FTid=VSP 005398.
FT CONFLICT 325 325 D -> G (IN REF. 1; BAB20277).
SQ SEQUENCE 455 AA; 49632 MW; 5CF31789C689AA CRC64;

Query Match 24.5%; Score 665.5; DB 1; Length 455;
Best Local Similarity 35.0%; Pred. No. 6.1e-42;
Matches 145; Conservative 66; Mismatches 176; Indels 27; Gaps 9;

QY 86 CFTLTGTFVLVGAALAAAGLLKFMGSKNSGIECDSSGTCINPNSWCDGVSHCPGGDE 145
Db 49 CVILGVLLAGAGIASMLLVLYLWPAASPS-----ISLTQEE-----EMTLNCEGVSC 99

QY 146 NRCV-----RLVGFPIQVYSSQKSHFVQDDWNNYGRACRDMGY-KNNFY 195
Db 100 EELLPSLPKTVSPRINGEDLLQVQVRAPDMLLVCHGWSPALGWHICKSLGHIRLTQH 159

QY 196 SSQGIVD--SGSTSFMKLNTSAGNVDIVKLYVSDACSKKAVSLRCIACGVNLNSRQ 253
Db 160 KAVNLSDIKLRQEAQLSARPGGL-VZEANKPANCPSPGRIVSLKCSBGAR---PLA 215

QY 254 SRVVGESALPGAMPQVSLHVQNVHVCSSITTEWIVTAACHVEK-PLANNPHWTATA 312
Db 216 SRVVGCAVASGRWPQASVMLGSRHTCGASVLAPHWVTAACHMYSFRLSRLSSRWVHA 275

QY 313 GILRQSFMYGAGYOVYKVIHPNDYKTKNDIALMKLOKPLTENDLVKVPNPGMW 372
Db 276 GLVSHGAVRCHQGTWVEKIIIPHLYSAQNHVDVALLQLRTPINPSDITDVAICLPAKEQY 335

QY 373 LQEQLCWISGWGATE-EKGKTSEVLNAKVLITQRCNSRYVDNLIITPAMICAGFLQ 431
Db 336 FPMGSCQWVSGWGHTDPSTHSSDTLQDTMVPVLLSTHLCNSCMYSGALTRMLCAGYLD 395

QY 432 GNVDSCQSGSGPLVTSKNNIWNWLGDTSGWGGCAKVRPGVYGVNMVFTDWIY 485
Db 396 GRADACQDSGGGLVCPSGDTWHLGVVSWGRCAPNRPVYAKVAEFLDWIH 449

RESULT 8
ENTK_BOVIN
ID ENTK_BOVIN STANDARD; PRT; 1035 AA.
AC P98072;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Enteroproteptidase precursor (EC 3.4.21.9) (Enterokinase).
GN P9857 OR ENTK.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]

```

```

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Ductenium;
RX MEDLINE=94329561; PubMed=8052624;
RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RL protease composed of a distinctive assortment of domains.";
RN Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94043122; PubMed=8226855;
RA Lavallie E.R., Rehmetulla A., Racie L.A., Dibiaseo E.A.,
RA Ferenc C., Grant K.L., Light A., McCoy J.M.;
RT "Cloning and functional expression of a cDNA encoding the catalytic
RN subunit of bovine enterokinase.";
RP J. Biol. Chem. 268:23311-23317(1993).
RP SEQUENCE OF 801-827.
RC TISSUE=Intestine;
RX MEDLINE=92189715; PubMed=1799406;
RA Light A., Janska H.;
RT "The amino-terminal sequence of the catalytic subunit of bovine
RN enterokinase.";
RL J. Protein Chem. 10:475-480(1991).
CC -1- FUNCTION: Responsible for initiating activation of pancreatic
CC proenzyme proenzymes (trypsin, chymotrypsin and carboxypeptidase
CC A). It catalyzes the conversion of trypsinogen to trypsin which in
CC turn activates other proenzymes including chymotrypsinogen,
CC procarboxypeptidases, and proelastases.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in
CC trypsinogen.
CC -1- SUBUNIT: Heterodimer of a catalytic (light) chain and a
CC multidomain (heavy) chain linked by a disulfide bond.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P98072-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P98072-2; Sequence=VSP 005386;
CC -1- TISSUE SPECIFICITY: Intestinal brush border.
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 1 MAM domain.
CC -1- SIMILARITY: Contains 1 SEA domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U09859; AAB40026.1; -
CC EMBL; L19663; AAA16035.1; -
CC FIR; A43090; A43090.
CC PDB; 1EXB; 14-OCT-99.
CC MEROPS; S01.156; -.
CC InterPro; IPR000859; CUB.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR002172; LDL_receptor_A.
CC InterPro; IPR000998; MAM domain.
CC InterPro; IPR001254; Peptidase S1.
CC InterPro; IPR001314; Peptidase SEA.
CC InterPro; IPR000882; SEA domain.
CC InterPro; IPR001190; Srcr_receptor.
CC Pfam; PF00431; CUB; 2.
CC Pfam; PF00657; ldl_recept_a; 2.
CC Pfam; PF00629; MAM; 1.

```

DR P1am; PFO1390; SEA; 1.
 DR P1am; PFO0530; SRCR; 1.
 DR P1am; PFO0089; CYP51; 1.
 DR PRINTS; PFO0722; CHYMOTRYPSIN.
 DR PRINTS; PFO0261; LDLRECEPTOR.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00192; LDLA; 2.
 DR SMART; SM00137; MAM; 1.
 DR SMART; SM00200; SEA; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00202; TYD; SPC; 1.
 DR PROSITE; PFO1180; CUB; 2.
 DR PROSITE; PFO1209; LDLRA; 1; 2.
 DR PROSITE; PFO068; LDLRA; 2; 2.
 DR PROSITE; PFO0740; MAM; 1; 1.
 DR PROSITE; PFO060; MAM; 2; 1.
 DR PROSITE; PFO024; SEA; 1.
 DR PROSITE; PFO0420; SRCR; 1; FALSE_NEG.
 DR PROSITE; PFO0287; SRCR; 2; 1.
 DR PROSITE; PFO0240; TRYPSIN_DOM; 1.
 DR PROSITE; PFO034; TRYPSIN_SER; 1.
 DR PROSITE; PFO0135; TRYPSIN_SER; 1.
 KW Signal-anchor; Glycoprotein; Myristate; Hydrolase; Serine protease;
 KW Zymogen; Transmembrane; Repeat; Alternative splicing; 3D-structure;
 KW Lipoprotein.

FT CHAIN 1 800 NON-CATALYTIC CHAIN (HEAVY CHAIN).
 FT CHAIN 801 1035 CATALYTIC CHAIN (LIGHT CHAIN).
 FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 48 1035 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 54 169 SEA.
 FT DOMAIN 197 238 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 240 350 CUB 1.
 FT DOMAIN 358 520 MAM.
 FT DOMAIN 540 650 MAM.
 FT DOMAIN 657 695 CUB 2.
 FT DOMAIN 694 787 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 801 1035 SRCR.
 FT DOMAIN 801 1035 SERINE PROTEASE.
 FT ACT_SITE 841 841 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 892 892 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 987 987 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT LIPID 2 2 N-myristoyl glycine (Potential).
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 206 225 BY SIMILARITY.
 FT DISULFID 219 236 BY SIMILARITY.
 FT DISULFID 659 671 BY SIMILARITY.
 FT DISULFID 666 684 BY SIMILARITY.
 FT DISULFID 678 693 BY SIMILARITY.
 FT DISULFID 788 912 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 826 842 BY SIMILARITY.
 FT DISULFID 926 933 BY SIMILARITY.
 FT DISULFID 957 972 BY SIMILARITY.
 FT DISULFID 983 1011 BY SIMILARITY.
 FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 263 263 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 486 486 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 698 698 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 722 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 762 762 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 864 864 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 903 903 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 166 192 Missing (in isoform Short).
 FT CONFLICT 808 808 /FTID=VSP_005386.
 SQ SEQUENCE 1035 AA; 114887 MW; E207970B08296E13 CRC64;
 Query Match 24.3%; Score 660; DB 1; Length 1035;
 Best Local Similarity 33.9%; Pred. No. 4.1e-41;
 Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;
 QY 65 VCTQPKSPSGTVCTSKTKKALCITLTI-----GFLVCAALAAAGLLWFKMGSKCSNSG 117
 DB 611 VTGPGPVNDVFSTNRTNMTVLFITDNLAKQGFANFTTGYGLG-----IPEPCKEEN 663
 QY 118 IECDSSTCTINFSNWCDSVSHCPGDEENRCVRLYG-----PNFILQVYSSQSKSWH 169
 DB 664 FQC-KDGEICPLVNLCDGFPCHKXGDSDEAHCVELFNGTTDSSGLVQFRIQ-----SIWH 716
 QY 170 PVOQDDNENYGRAACRDMGYKNFYSSQGVIVDSSTGTFMKLNTSAGNVDIYKLYHSD 229
 DB 717 VACAENMTQISDVBVQLGLGFG--NSVPTFTSTGGPYVNLNTAPNSLI---LTPSQ 771
 QY 230 ACSSKAVVSLRC--IACGVNLSRSQS-RIVGGESALPGAWPQVSLHVQNVHVCSSII 286
 DB 772 QCLEDSLILLQCNVYKSCGKKLVQEVSPKIVGSDSREGAWPWVALYFDDQDQVCGASLV 831
 QY 287 TPEWIVTAACVEXPLANNPHWTAFAGILARQSPMFYAGYQVE-----KVISHPNYDSKT 341
 DB 832 SRDLVLSAAHCYVGRNNEPSKKAVALGLNWSNL---TSPQIEIETRLIDQIVINPHYNKR 888
 QY 342 KNDIALMKLQKPLTFNDLVKVPCLPNPMMLOPEOLCWSIGSWGATEEKGKTSVLNAAK 401
 DB 889 KNDIAMHLEMKVNYTDYIOPICLPEENQVFPFGPICSIAAGWALIYQGSTADVLQEAD 948
 QY 402 VLLIETORCSNRYVYDNLITPAMICAGFLOGNVDSCGSGGGLVTSKKNWILGDTSW 461
 DB 949 VPLLSNEKQCCQOMPEYN-ITENMVCAGYEAGGVDSQSGSGGLMCQENRMLLAGVTSF 1007
 QY 462 GSGCAKAYRPGVYGNVMVFTDWI 484
 DB 1008 GYQCALPNRPGVYARVPRFTEWI 1030
 RESULT 9
 ID TM55 HUMAN STANDARD; PRT; 457 AA.
 AC Q9H3S3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
 GN TMFRSS5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain.
 RX PubMed=11741866;
 RA Yamaguchi N., Okui A., Yamada T., Nakazato H., Mitsui S.; cloned from
 human spinal cord.;
 RT "Spinesin/TMPRSS5, a novel transmembrane serine protease, expressed in
 CC -|- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
 CC -|- TISSUE SPECIFICITY: Brain-specific. Predominantly expressed in
 CC neurons, in their axons, and at the synapses of motoneurons in the
 CC spinal cord.
 CC -|- SIMILARITY: Belongs to peptidase family S1.
 CC -|- SIMILARITY: Contains 1 SRCR domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see [http://www.isb-sib.ch/announce/](http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch) or send an email to license@isb-sib.ch).

EMBL; A8028140;	BAB20375.1;	-
DR	HSP; P00763;	LDPO.
DR	Genew; HGNC:14908;	TMPRSS5.
DR	MIM; 606751;	-
DR	MEROPS; S01.313;	-
DR	InterPro; IPR009003;	Cys_Ser_trypsin.
DR	InterPro; IPR001254;	Peptidase_S1.
DR	InterPro; IPR001314;	Peptidase_S1A.
DR	InterPro; IPR001190;	Ser_rceptor.
DR	Pfam; PF00089;	trypsin; 1.
DR	PRINTS; PR00722;	CHYMOTRYPSIN.
DR	SMART; SM00020;	TyP_Spc; 1.
DR	PROSITE; PS50240;	TRYPSIN_DOM; 1.
DR	PROSITE; PS00134;	TRYPSIN_HIS; 1.
DR	PROSITE; PS00135;	TRYPSIN_SER; 1.
DR	PROSITE; PS00420;	SRCR_1; FALSE NEG.
DR	PROSITE; PS0287;	SRCR_2; FALSE NEG.
KW	Hydrolase;	Serine protease; Transmembrane; Signal-anchor;
KW	Glycoprotein.	1
FT	DOMAIN	1 49 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
FT	DOMAIN	71 457 EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	112 207 SRCL.
FT	DOMAIN	218 457 SERINE PROTEASE.
FT	ACT_SITE	258 258 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	217 218 CLEAVAGE (POTENTIAL).
FT	DLSULFID	135 196 BY SIMILARITY.
FT	DLSULFID	148 205 BY SIMILARITY.
FT	DLSULFID	209 328 BY SIMILARITY.
FT	DLSULFID	243 259 BY SIMILARITY.
FT	DLSULFID	374 390 BY SIMILARITY.
FT	DLSULFID	401 429 BY SIMILARITY.
FT	CARBHYD	163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBHYD	170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBHYD	195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBHYD	319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBHYD	375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQ	SEQUENCE	457 AA; 449574 MW; 64406AB498A2651 CRC64;

Query Match	24.0%;	Score	651.5;	DB 1;	Length	457;			
Best Local Similarity	32.4%;	Pred. No.	6.7e-41;						
Matches	161;	Conservative	63;	Mismatches	194;	Indels	83;	Gaps	16;

QY	8	PPAIGPYENHGYQENPYPAQPTVVTVPVVEHPA-QYPSFPVQVAPRVLTQASNPVVC	66
DB	9	PMMEAGYAE-----EGPGPG-----IPRAFPGDQHP-----ISQA-----	39
QY	67	TQPKPSGTCTSKTKKALCITLTGLTFGLVGAALAGLLWKFMGSKCSN--SGTECDSSG	124
DB	40	-----VCWESMRGCAVLGALG-LLAGAGVGSWLLVLLCPAASQPISGTLQDEEI	89
QY	125	TCINPSNWCQGVSHCPGGBENRCV-----RLYGPRIILQVYSSQKSNWHPVCODD	175
DB	90	TL-----SCSEASEALLPALPTVFRINSEDFLEAQVDRQPRMLLVCHEG	138
QY	176	WNEYNGRAACRDMGY-KNNFYSSOGIVDDSGSTFMKLNTSAGNVDIYKKL--YHSDA--	230
DB	139	WSPALGLQICWSLGHLLRTHHKGWNLTD-----IKLNSQEPALSPRLGPGFLEAWQ	191
QY	231	-----CSSKAVYSLECIACGVNLNSSROSRIVGESALPGAWPMQVSLHVQNHVHVCGSII	286
DB	192	FRNNCTSGQVVSLRCSCEGAR--PLASIRVGGQSVAPGRWPMQASVALGRHTCCGSVL	248
QY	287	TPEWIVTAAHCVEK-PLNNPWHMTAFAGILRQSPFMFYGAGYQVEKVIISHPNYDSKTKNND	345

[illegible]

turn activates other proenzymes including chymotrypsinogen, procarboxypeptidases, and proelastases.
-!- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in trypsinogen.
-!- SUBUNIT: Heterodimer of a catalytic (light) chain and a multidomain (heavy) chain linked by a disulfide bond.
-!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
-!- TISSUE SPECIFICITY: Intestinal brush border.
-!- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS CLEAVED BY A TRYPSIN-LIKE PROTEASE.
-!- DISEASE: Defects in PRSS7 are a cause of enterokinase deficiency [MIM:246200]; a life-threatening intestinal malabsorption disorder characterized by diarrhea and failure to thrive.
-!- SIMILARITY: Belongs to peptidase family S1.
-!- SIMILARITY: Contains 2 CUB domains.
-!- SIMILARITY: Contains 1 LDL-receptor class A domains.
-!- SIMILARITY: Contains 1 SEA domain.
-!- SIMILARITY: Contains 1 SRCR domain.
-!- SIMILARITY: Contains 1 SRCR domain.
This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
EMBL; U09860; AAC50138.1; -
EMBL; Y19124; CAB65555.1; -
EMBL; Y19125; CAB65555.1; JOINED.
EMBL; Y19126; CAB65555.1; JOINED.
EMBL; Y19127; CAB65555.1; JOINED.
EMBL; Y19128; CAB65555.1; JOINED.
EMBL; Y19129; CAB65555.1; JOINED.
EMBL; Y19130; CAB65555.1; JOINED.
EMBL; Y19131; CAB65555.1; JOINED.
EMBL; Y19132; CAB65555.1; JOINED.
EMBL; Y19133; CAB65555.1; JOINED.
EMBL; Y19134; CAB65555.1; JOINED.
EMBL; Y19135; CAB65555.1; JOINED.
EMBL; Y19136; CAB65555.1; JOINED.
EMBL; Y19137; CAB65555.1; JOINED.
EMBL; Y19138; CAB65555.1; JOINED.
EMBL; Y19139; CAB65555.1; JOINED.
EMBL; Y19140; CAB65555.1; JOINED.
EMBL; Y19141; CAB65555.1; JOINED.
EMBL; Y19142; CAB65555.1; JOINED.
EMBL; Y19143; CAB65555.1; JOINED.
EMBL; AL163218; CAB90392.1; -
EMBL; AL163217; CAB90389.1; -
PIR; A56318; A56318.
HSP; P00783; LDPO.
MEROPS; S01.156; -
Genew; HGNC:9490; PRSS7.
MIM; 606635; -
MIM; 246200; -
GO; GO:0005903; C:brush border; TAS.
InterPro; IPR000859; CUB.
InterPro; IPR009003; Cys_Ser_trypsin.
InterPro; IPR002172; LDL_receptor_A.
InterPro; IPR000998; MAM domain.
InterPro; IPR001254; Peptidase_S1.
InterPro; IPR001314; Peptidase_S1A.
InterPro; IPR000082; SEA domain.
InterPro; IPR001190; Srcr_receptor.
Pfam; PF00431; CUB; 2.
Pfam; PF00057; ldl_recept_a; 2.
Pfam; PF00829; MAM; 1.
Pfam; PF01390; SEA; 1.
Pfam; PF00530; SRCR; 1.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.

PRINTS; PR00261; LDLRECEPTOR.
PRINTS; PR00020; MAMDOMAIN.
SMART; SM00042; CUB; 2.
SMART; SM00192; LDLa; 2.
SMART; SM00137; MAM; 1.
SMART; SM00200; SEA; 1.
SMART; SM00202; SR; 1.
SMART; SM00020; Tryp_SPC; 1.
PROSITE; PS01180; CUB; 2.
PROSITE; PS01209; LDLa; 1; 2.
PROSITE; PS00068; LDLa; 2; 2.
PROSITE; PS00740; MAM; 1; 1.
PROSITE; PS00060; MAM; 2; 1.
PROSITE; PS00024; SEA; 1.
PROSITE; PS00420; SRCR; 1; FALSE_NEG.
PROSITE; PS0287; SRCR; 2; 1.
PROSITE; PS0240; TRYPSIN_DOM; 1.
PROSITE; PS0134; TRYPSIN_HIS; 1.
PROSITE; PS00135; TRYPSIN_SER; 1.
KW Signal-anchor; Glycoprotein; Myristate; Hydrolase; Serine protease; Zymogen; Transmembrane; Repeat; Lipoprotein.
FT CHAIN 1 784 NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT CHAIN 785 1019 CATALYTIC CHAIN (LIGHT CHAIN).
FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
FT DOMAIN 48 1019 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 52 169 SEA.
FT DOMAIN 182 223 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 225 334 CUB 1.
FT DOMAIN 342 504 MAM.
FT DOMAIN 524 634 CUB 2.
FT DOMAIN 641 679 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 678 771 SRCR.
FT DOMAIN 785 1019 SERINE PROTEASE.
FT ACT_SITE 825 825 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 876 876 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 971 971 N-myristoyl glycine (Potential).
FT LIPID 2 2 BY SIMILARITY.
FT DISULFID 184 197 BY SIMILARITY.
FT DISULFID 191 210 BY SIMILARITY.
FT DISULFID 204 221 BY SIMILARITY.
FT DISULFID 643 655 BY SIMILARITY.
FT DISULFID 650 668 BY SIMILARITY.
FT DISULFID 662 677 BY SIMILARITY.
FT DISULFID 772 896 INTERCHAIN (BY SIMILARITY).
FT DISULFID 810 826 BY SIMILARITY.
FT DISULFID 910 977 BY SIMILARITY.
FT DISULFID 941 956 BY SIMILARITY.
FT DISULFID 967 995 BY SIMILARITY.
FT CARBOHYD 116 116 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 328 328 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 335 335 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 388 388 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 440 440 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 503 503 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 630 630 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 682 682 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 706 706 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 848 848 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 887 887 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 909 909 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 949 949 N-LINKED (GLCNAC... (POTENTIAL).
FT CONFLICT 134 134 Q -> E (IN REF. 3).
FT CONFLICT 732 732 S -> P (IN REF. 3).
FT CONFLICT 754 771 SQGLQDSLRQCNHKS -> RNNANEIDALSPILLIA (IN REF. 3).


```

Query Match      23.8%; Score 647; DB 1; Length 1019;
Best Local Similarity 35.8%; Pred. No. 3.7e-40;
Matches 138; Conservative

QY 113 CSNSGICDSCTGCTNPNSWCDGSHCPGEGDENRCVRLYG-----PNFILQVSSQ 164
DB 543 CKADHFC-KNGECVPLVNICDGLHCEGDSBACVRFNGTNNGLVRFRIQ-----696
QY 165 RKSHPYQCDNMENYGRACRDMGKYNFYSSQIGVDDSGSTFWMKLTSAAGNDVIYK 224
DB 697 -SIWHTACAEWTQTQISNDVQCQLLGLSG-NSSKPIFSTDGG-PFVKLTAP---DGLH 750
QY 225 LYHSDACSSXAVSLRC--TACGVNLNSR-QSRIVGGSALPGAPWQVSHVQNVHVC 281
DB 751 LTPSQQLQSLRLQCNHKSCKGLAODITPKIVGSSNAKEGAPWVGLYGGRLLC 810
QY 282 GSGIITPEWITVTAACHVEKPLNPNWHTAFAGILRQSFNF--YGAGYQVEKVIHPNYDS 339
DB 811 GASLVSSDMLVSAACHVYGRNLEPSKWTALGLEMSNLTSPTQVPRLDIEIVNPHYR 870
QY 340 KTKNDIATMKLQKPLTFNDLVKPVCLPNPGLMLOPEQLCWTSWGATEEKGKTSVLNA 399
DB 871 RKNNDIAMHLEFKNYTYDIIQICLPENQVFPFPRNCSTAGWTVVYQGTANILQE 930
QY 400 AKVLLIETORCNSRYVVDNLITPAMICAGFLOGVDSQCGSGGPLVTSKNINWLGDT 459
DB 931 ADVPLLNSERCQQQMEYN-ITENMICAGYEGLDSCQDSGGPLMCCENRFLAGVT 989
QY 460 SWGSGCKAVRPGVGNVWFTDWI 484
DB 990 SFGYKCALPNRPGVYARVERFTEWI 1014

RESULT 11
ENTK_MOUSE
ID_ENTK_MOUSE STANDARD; PRT: 1069 AA.
AC P97435;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Enteropeptidase (EC 3.4.21.9) (Enterokinase).
GN PRS7 OR ENTK
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Ductenum;
RX MEDLINE=98147142; PubMed=9456188;
RA Yuan X, Zheng X, Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.;
RT "Structure of murine enterokinase (enteropeptidase) and expression in
RT small intestine during development.";
RL Am. J. Physiol. 274:G342-G349(1998).
CC -1- FUNCTION: Responsible for initiating activation of pancreatic
CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
CC A). It catalyzes the conversion of trypsinogen to trypsin which in
CC turn activates other proenzymes including chymotrypsinogen,
CC procarboxypeptidases, and proelastases (By similarity).
CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-|-Ile-7 bond in
CC trypsinogen.
CC -1- SUBUNIT: Heterodimer of a catalytic (light) chain and a
CC multidomain (heavy) chain linked by a disulfide bond (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 1 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 1 NAM domain.
CC -1- SIMILARITY: Contains 1 SEA domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.

```

```

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: U73378; AAB37317.1; -.
HSP: Q07954; ICR8.
MEROPS: S01.156; -.
MGD: MGI:1197523; Pres7.
InterPro: IPR000859; CUB.
InterPro: IPR009003; Cys Ser trypsin.
InterPro: IPR002172; LDL_receptor_A.
InterPro: IPR000998; MAM domain.
InterPro: IPR001254; Peptidase_S1.
InterPro: IPR001314; Peptidase_S1A.
InterPro: IPR009020; Protease_inhib.
InterPro: IPR000082; SEA domain.
DR InterPro: IPR001190; Srcf_receptor.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00057; ldl_recept_a; 2.
DR Pfam: PF00629; MAM; 1.
DR Pfam: PF01390; SEA; 1.
DR Pfam: PF00530; SRCR; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PRINTS: PR00020; MAMDOMAIN.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00192; LDLA; 2.
DR SMART: SM00137; MAM; 1.
DR SMART: SM00200; SEA; 1.
DR SMART: SM00020; SR; 1.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01209; LDLRA_1; 2.
DR PROSITE: PS00668; LDLRA_2; 2.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS00660; MAM_2; 1.
DR PROSITE: PS00024; SEA; 1.
DR PROSITE: PS00420; SRCR; 1.
DR PROSITE: PS00287; SRCR_2; 1.
DR PROSITE: PS00340; TRYP_SIN_DOM; 1.
DR PROSITE: PS00134; TRYP_SIN_SER; 1.
DR PROSITE: PS00135; TRYP_SIN_SER; 1.
KW Signal-anchor; Glycoprotein; Hydrolase; Serine protease; Zymogen;
KW Transmembrane; Repeat.
FT CHAIN 1 829
FT CHAIN 830 1069
FT TRANSMEM 19 47
FT DOMAIN 48 1069
FT DOMAIN 52 169
FT DOMAIN 227 268
FT DOMAIN 270 379
FT DOMAIN 387 549
FT DOMAIN 569 679
FT DOMAIN 686 724
FT DOMAIN 723 816
FT DOMAIN 730 1069
FT ACT_SITE 874 874
FT ACT_SITE 925 925
FT ACT_SITE 1021 1021
FT DISULFID 229 242
FT DISULFID 236 255
FT DISULFID 249 266
FT DISULFID 688 700
FT DISULFID 695 713
FT DISULFID 707 722
NON-CATALYTIC CHAIN (HEAVY CHAIN).
CATALYTIC CHAIN (LIGHT CHAIN).
CYTOPLASMIC (POTENTIAL).
SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
EXTRACELLULAR (POTENTIAL).
SEA.
LDL-RECEPTOR CLASS A 1.
CUB 1.
MAM.
CUB 2.
LDL-RECEPTOR CLASS A 2.
SRCR.
SERINE PROTEASE.
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.

```

FT DISULFID 817 945 INTERCHAIN (BY SIMILARITY).
FT DISULFID 859 975 BY SIMILARITY.
FT DISULFID 959 1027 BY SIMILARITY.
FT DISULFID 991 1006 BY SIMILARITY.
FT DISULFID 1017 1045 BY SIMILARITY.
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 579 579 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 675 675 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 727 727 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 751 751 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 770 770 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 791 791 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 936 936 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 999 999 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1069 AA; 118735 MW; 562549E463743C3D CRC64;

Query Match. 23.5%; Score 638; DB 1; Length 1069;
Best Local Similarity 30.5%; Pred No. 1.8e-39;
Matches 164; Conservative 92; Mismatches 196; Indels 86; Gaps 21;

QY 18 HGQPNPYPAPQTVVPT-----VVEV-HPAQYPSVPVQVAPRVLTQAS---- 61
Db 544 NGICQSQYP-EPTLVTPPELPDTCGPFELMEPNSTFSS--PNFPDKYPNQASICWN 600
QY 62 -----NPV-----VCTQPKSPSGTCTKTKAL 85
Db 601 LNAQRGKNTQLHFQEDLENINDVVEVDGGEFFSLILAVTGP-GPVKDLFSITNMTV 659
QY 86 CHTLGLTFLVG--AALAAGLWKPMG--SKCSNGIECDSSGTCINPSNWCDSVHCPG 141
Db 660 IFTTNMETRRKFKANFTSGY---YLGPEPCQDDEFQC-KDGNCPILGNCLCDSYPRCD 715
QY 142 GEDENRCVRLY---GPNFLOVYSSQRKSHVPVCCDDWENYGRAACRDMGYKNFYSS 197
Db 716 GSDEASCVRFLNGTSNNGLVQF--NIHSIWHIACAENWTTQISNEVCHLLGLGS--ANS 771
QY 198 QGIVDSDGTSFMKNTSAGNVDIYKKLYHSDACSSKAVSLRC--IACGVNLSRSQS- 254
Db 772 SMPISSTGGGPFVRVQNAQNGSLI---LTPSLQCSQDSLILLQCNHKSCKEKKVTQKVP 828
QY 255 RIUGGESALPGAPWQVSLHVNVRH---VCGSGTITPEWITVTAHCKVEKPLNKPWHTA 310
Db 829 KIVGSDAQAQAPWVVALYHEDRSTDRLLCQSLVSSDWLVSAHCVYRNLDPTRTWA 888
QY 311 FAGILRQSPFYGAGYQ--VEKVISHPNYDSKTRNDIALMKLQKPLTFNDLVKPVCLPN 368
Db 889 VLGLHMQSNLTSPQVVRVVDQIVINPHYDRRRKYNVDIAMWHLEFKVNYTDYIQICLPE 948
QY 369 PGWMLQPEQLCWISGWGATE-EKGKTSVLNAAKVLLETQKNSRYVYDNLITPAMICA 427
Db 949 ENQIFPGTCTCIAGCYDKINAGSTVDVLKADVPLISNEKCCQQLPEYN-ITESMICA 1007
QY 428 GFLOQNVDSQDGSGLPVTSKNTLWILGTSWGSCKAKAYRPGVGVNVMVFTDWTY 485
Db 1008 GYEEGGIDSCQDGSGLPQCMENNRWFLVGVTSFGVQCALPNHPGVYRVVSQFIENIWH 1065

RESULT 12
ENTK_PIG
ID_ENTK_PIG STANDARD; PRT; 1034 AA.
AC P98074;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Enteropeptidase precursor (BC 3.4.21.9) (Enterokinase).
GN PRSS7 OR ENTK.

Sus scrofa (Pig).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
[1] NCBI_TaxId=9823;
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
TISSUE=Ductal mucosa;
MEDLINE=94327548; PubMed=8051081;
Matsushima M., Ichinose M., Yanagi N., Kakei N., Tsukada S.,
Miki K., Kurokawa K., Tashiro K., Shiohara K., Shinomiya K.,
Uneyama H., Inoue H., Takahashi T., Takahashi K.,
"Structural characterization of porcine enteropeptidase.";
J. Biol. Chem. 269:19976-19982(1994).
CC -!- FUNCTION: Responsible for initiating activation of pancreatic
proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
A). It catalyzes the conversion of trypsinogen to trypsin which in
turn activates other proenzymes including chymotrypsinogen,
procarboxypeptidases, and proelastases.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-|-Ile-7 bond in
trypsinogen.
CC -!- SUBUNIT: Heterotrimer of a catalytic (light) chain, a multidomain
(heavy) chain, and a mini chain.
CC -!- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 1 MAM domain.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC -!- SIMILARITY: Contains 1 SRCR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; D30799; BAB06459.1; -.
CC HSP; P00763; IDPO.
CC MEROPS; S01.156; -.
CC InterPro; IPR000859; CUB.
CC InterPro; IPR009003; Cys Ser trypsin.
CC InterPro; IPR002172; LDL_receptor_A.
CC InterPro; IPR000998; MAM_domain.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC InterPro; IPR000082; SEA_domain.
CC InterPro; IPR001190; Srcr_receptor.
CC Pfam; PF00431; CUB; 2.
CC Pfam; PF00629; MAM; 1.
CC Pfam; PF01390; SEA; 1.
CC Pfam; PF00530; SRCR; 1.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PRINTS; PR00261; LDLRECEPTOR.
CC PRINTS; PR00020; MAMDOMAIN.
CC SMART; SM00042; CUB; 2.
CC SMART; SM00192; LDLA; 2.
CC SMART; SM00137; MAM; 1.
CC SMART; SM00200; SEA; 1.
CC SMART; SM00202; SR; 1.
CC SMART; SM00020; Tryp_Spc; 1.
CC PROSITE; PS01180; CUB; 2.
CC PROSITE; PS01209; LDLRA_1; 2.
CC PROSITE; PS50068; LDLRA_2; 2.
CC PROSITE; PS00740; MAM_1; 1.
CC PROSITE; PS50060; MAM_2; 1.
CC PROSITE; PS50024; SEA; 1.

176 WNEYGRAACRDMGYKQK-----FYSSQGVDDSGSTSPMKLNTSAGNVDIYKLYHSDA 230
722 WTTQTSDDVYQQLGLGTGNSSMPFFS-----SGGPEFKLNTAPNGSLI-----LTASEQ 771
231 CSKAVVSLRC--IAGVNLNRSROS-RIVGESALPGAWPQVSLVQNVHVCGGSIIT 287
772 CREDSILLQCHKSCGKKQVQAEVSPKIVGNDSEGAWPVVALYNGQLLGGASLVS 831
288 PEWITAAHCVEKPLNPNHWTAFAGILRQSFMYFVAGYQ-----VEKVISHPNYDSKTK 342
832 RDLVLSAAHCVTVGRNLEPSPKAILGLHMTSNL---TSPQIVTRLIDEIVINPHYNNRRK 888
343 NNDIALMKLQKPLTFNDLVKPCVCLPNPGLMLOPEOLCWISGMGATBEKGTSEVLNAKV 402
889 DSDIAWHLEFPKVNVDYIQICLPEENQVFPFGICSIAGKVIYQSPADILQEADV 948
403 LIETQRCNSRYVDNLITPAMICAGFLQGVDSQCGSGPLVTSKNNIWLIGTWSG 462
949 PLLSNEKCOQOMPEYN-ITENMCAGYEKGIDSCQDSCGSLMCLNRRLLAGVTSEF 1007
463 SGCAKAYRPGVYGVNVVFTDWI 484
1008 YQCALPFRFVGVARVPKFTWI 1029
RESULT 13
CORI_HUMAN STANDARD; PRT; 1042 AA.
ID CORI_HUMAN STANDARD; PRT; 1042 AA.
AC Q9Y5Q5; Q9UHY2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Atrial natriuretic peptide-converting enzyme (EC 3.4.21.-) (pro-ANP-
converting enzyme) (Corin) (Heart specific serine proteinase ATC2).
GN CRN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=99262646; PubMed=10329693;
RA Yan W., Sheng N., Seto M., Morser J., Wu Q.;
RT "Corin, a mosaic transmembrane serine protease encoded by a novel cDNA
from human heart";
RL J. Biol. Chem. 274:14926-14935 (1999).
RN [2]
RP SEQUENCE OF 734-1040 FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=20534769; PubMed=11082206;
RA Hooper J.D., Searman A.L., Clarke B.E., Normyle J.F., Antalis T.M.;
RT "Localization of the mosaic transmembrane serine protease corin to
heart myocytes";
RL Eur. J. Biochem. 267:6931-6937 (2000).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=20359740; PubMed=10880574;
RA Yan W., Wu F., Morser J., Wu Q.;
RT "Corin, a transmembrane cardiac serine protease, acts as a pro-atrial
natriuretic peptide-converting enzyme";
RL Proc. Natl. Acad. Sci. U.S.A. 97:8525-8529 (2000).
CC -!- FUNCTION: Converts Pro-ANP to ANP. Cleaves Pro-ANP specifically
between Arg-123 and Ser-124.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC -!- TISSUE SPECIFICITY: Highly expressed in heart. Expressed in heart
myocytes.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 2 frizzled (FZ) domains.
CC -!- SIMILARITY: Contains 7 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 1 SCRD domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS0134; TRYPSIN_HIS; 1.
DR PROSITE; PS0135; TRYPSIN_SER; 1.
KW Signal-anchor; Glycoprotein; Repeat; Myristate; Hydrolase; Serine protease;
KW Zymogen; Transmembrane; Repeat; Lipoprotein.
FT CHAIN 52 117
FT CHAIN 118 799
FT CHAIN 800 1034
FT DOMAIN 1 18
FT TRANSMEM 19 47
FT DOMAIN 48 1034
FT DOMAIN 52 169
FT DOMAIN 197 238
FT DOMAIN 240 349
FT DOMAIN 357 519
FT DOMAIN 539 649
FT DOMAIN 656 694
FT DOMAIN 693 786
FT DOMAIN 800 1034
FT ACT_SITE 840 840
FT ACT_SITE 891 891
FT ACT_SITE 986 986
FT LIPID 2 2
FT DISULFID 199 212
FT DISULFID 206 225
FT DISULFID 219 236
FT DISULFID 658 670
FT DISULFID 665 683
FT DISULFID 677 692
FT DISULFID 787 911
FT DISULFID 835 841
FT DISULFID 925 992
FT DISULFID 956 971
FT DISULFID 982 1010
FT CARBOHYD 116 116
FT CARBOHYD 147 147
FT CARBOHYD 170 170
FT CARBOHYD 194 194
FT CARBOHYD 283 283
FT CARBOHYD 343 343
FT CARBOHYD 350 350
FT CARBOHYD 403 403
FT CARBOHYD 455 455
FT CARBOHYD 485 485
FT CARBOHYD 518 518
FT CARBOHYD 549 549
FT CARBOHYD 645 645
FT CARBOHYD 697 697
FT CARBOHYD 701 701
FT CARBOHYD 721 721
FT CARBOHYD 740 740
FT CARBOHYD 761 761
FT CARBOHYD 804 804
FT CARBOHYD 863 863
FT CARBOHYD 902 902
FT CARBOHYD 964 964
FT SEQUENCE 1034 AA; 114776 MW; 038C64CF64CC368 CRC64;
Query Match 23.3%; Score 634; DB 1; Length 1034;
Best Local Similarity 33.3%; Pred. No. 3.4e-39;
Matches 147; Conservative 72; Mismatches 179; Indels 44; Gaps 13;
QY 65 VCTQKSPSGTCTKTKALCIT---LTLG---TELVGALAGLWKGKSCNSG 117
DB 610 VYTGPGVEDVFTNRMVTLFITNDALTKGFKANFTGVHLG-----IPECKEDN 662
QY 118 IECDSGTCINPSNWCDSVSHCPGGEDENRCVRLYG--PNFTLQYSSQRKSWHPVCDD 175
DB 663 FQCE-NGECVLLVNLCDGFSHCKDGSDEAHCVRFLNGTANNGLVQFRIQSIWHTACAEN 721

disease.";
 Genomics 69:225-234 (2000).
 [3]
 RP SEQUENCE FROM N.A., AND VARIANTS SER-143; THR-178; GLN-202; CYS-269;
 VAL-311; ALA-358; ALA-381; PRO-442 AND GLN-560.
 RA Rieder M.J., Arnel T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,
 RA Rajkumar N., Toth E.J., Yi Q., Nickerson D.A.,
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBSJ databases.
 [4]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE-91152016; PubMed-198666;
 RA McMullen B.A., Fujikawa K., Davie E.W.;
 RT "Location of the disulfide bonds in human plasma prekallikrein: the
 RT presence of four novel apple domains in the amino-terminal portion of
 RT the molecule.";
 RL Biochemistry 30:2050-2056 (1991).
 [5]
 RP CARBOHYDRATE-LINKAGE SITE ASN-453.
 RX MEDLINE-22660472; PubMed-12754519;
 RA Zhang H., Li X.-J., Martin D.B., Aebersold R.;
 RT "Identification and quantification of N-linked glycoproteins using
 RT hydrazide chemistry, stable isotope labeling and mass spectrometry.";
 RL Nat. Biotechnol. 21:660-666 (2003).
 CC -1- FUNCTION: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It
 CC activates, in a reciprocal reaction, factor XII after its binding
 CC to a negatively charged surface. It also releases bradykinin from
 CC HMW kininogen and may also play a role in the renin-angiotensin
 CC system by converting prorenin into renin.
 CC -1- CATALYTIC ACTIVITY: Cleaves selectively Arg-|-Xaa and Lys-|-Xaa
 CC bonds, including Lys-|-Arg and Arg-|-Ser bonds in (human)
 CC kininogen to release bradykinin.
 CC -1- SUBUNIT: The zymogen is activated by factor XIIa, which cleaves
 CC the molecule into a light chain, which contains the active site,
 CC and a heavy chain, which associates with HMW kininogen. These
 CC chains are linked by one or more disulfide bonds.
 CC -1- DISEASE: Defects in KLKB1 are the cause of Fletcher factor
 CC deficiency (MIM:229000); a blood coagulation defect.
 CC -1- SIMILARITY: Belongs to peptidase family S1. Plasma kallikrein
 CC subfamily.
 CC -1- SIMILARITY: Contains 4 apple domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M13143; AAA60153.1; .
 DR EMBL; AF232742; AAF79940.1; JOINED.
 DR EMBL; AF232734; AAF79940.1; JOINED.
 DR EMBL; AF232735; AAF79940.1; JOINED.
 DR EMBL; AF232736; AAF79940.1; JOINED.
 DR EMBL; AF232737; AAF79940.1; JOINED.
 DR EMBL; AF232738; AAF79940.1; JOINED.
 DR EMBL; AF232739; AAF79940.1; JOINED.
 DR EMBL; AF232740; AAF79940.1; JOINED.
 DR EMBL; AF232741; AAF79940.1; JOINED.
 DR EMBL; AY190920; AAN84794.1; .
 DR PIR; A00921; KQHUP.
 DR HSSP; P00763; LDPO.
 DR MEROPS; S01.212; .
 DR Genew; HGNC:6371; KLKB1.
 DR MIM; 229000; .
 DR GO; GO:0003807; F1plasma kallikrein activity; TAS.
 DR GO; GO:0006508; P1proteolysis and peptidolysis; TAS.
 DR InterPro; IPR000177; Apple.
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR003014; PAN.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001314; Peptidase_S1a.
 DR Pfam; PF00024; PAN; 4.

DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0005; APPELDDOMAIN.
 DR PRINTS; PRO0722; CHYMOTRYPSIN.
 DR SMART; SMO0223; APPLE; 4.
 DR SMART; SMO0020; Tryp_Spc; 1.
 DR PROSITE; PS00495; APPLE; 4.
 DR PROSITE; PS02040; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
 KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
 KW Repeat; Polymorphism.
 FT SIGNAL 1 19
 FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
 FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
 FT DOMAIN 20 105 APPLE 1.
 FT DOMAIN 110 195 APPLE 2.
 FT DOMAIN 200 285 APPLE 3.
 FT DOMAIN 291 376 APPLE 4.
 FT DOMAIN 389 621 SERINE PROTEASE.
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .).
 FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
 FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
 FT DISULFID 21 104 CHARGE RELAY SYSTEM.
 FT DISULFID 47 77
 FT DISULFID 51 57
 FT DISULFID 111 194
 FT DISULFID 137 166
 FT DISULFID 141 147
 FT DISULFID 201 284
 FT DISULFID 227 256
 FT DISULFID 231 237
 FT DISULFID 292 375
 FT DISULFID 318 347
 FT DISULFID 322 328
 FT DISULFID 340 345
 FT DISULFID 383 503
 FT DISULFID 419 435
 FT DISULFID 517 584
 FT DISULFID 548 563
 FT DISULFID 574 602
 FT VARIANT 143 143
 FT VARIANT 178 178
 FT VARIANT 202 202
 FT VARIANT 208 208
 FT VARIANT 269 269
 FT VARIANT 311 311
 FT VARIANT 358 358
 FT VARIANT 381 381
 FT VARIANT 442 442
 FT VARIANT 560 560
 FT SEQUENCE 638 AA; 71369 MW; E62F9C1053838FB4 CRC64;
 N -> S (common polymorphism).
 A -> T.
 H -> Q.
 H -> P.
 S -> C.
 F -> V.
 T -> A.
 S -> A.
 Q -> P.
 R -> Q.
 /FTID=VAR_013598.
 /FTID=VAR_016280.
 /FTID=VAR_013599.
 /FTID=VAR_013600.
 /FTID=VAR_016281.
 /FTID=VAR_016282.
 /FTID=VAR_016283.
 /FTID=VAR_016284.
 /FTID=VAR_016285.
 /FTID=VAR_016286.
 /FTID=VAR_016287.
 /FTID=VAR_016288.
 /FTID=VAR_016289.
 /FTID=VAR_016290.
 /FTID=VAR_016291.
 /FTID=VAR_016292.
 /FTID=VAR_016293.
 /FTID=VAR_016294.
 /FTID=VAR_016295.
 /FTID=VAR_016296.
 /FTID=VAR_016297.
 /FTID=VAR_016298.
 /FTID=VAR_016299.
 /FTID=VAR_016300.
 /FTID=VAR_016301.
 /FTID=VAR_016302.
 /FTID=VAR_016303.
 /FTID=VAR_016304.
 /FTID=VAR_016305.
 /FTID=VAR_016306.
 /FTID=VAR_016307.
 /FTID=VAR_016308.
 /FTID=VAR_016309.
 /FTID=VAR_016310.
 /FTID=VAR_016311.
 /FTID=VAR_016312.
 /FTID=VAR_016313.
 /FTID=VAR_016314.
 /FTID=VAR_016315.
 /FTID=VAR_016316.
 /FTID=VAR_016317.
 /FTID=VAR_016318.
 /FTID=VAR_016319.
 /FTID=VAR_016320.
 /FTID=VAR_016321.
 /FTID=VAR_016322.
 /FTID=VAR_016323.
 /FTID=VAR_016324.
 /FTID=VAR_016325.
 /FTID=VAR_016326.
 /FTID=VAR_016327.
 /FTID=VAR_016328.
 /FTID=VAR_016329.
 /FTID=VAR_016330.
 /FTID=VAR_016331.
 /FTID=VAR_016332.
 /FTID=VAR_016333.
 /FTID=VAR_016334.
 /FTID=VAR_016335.
 /FTID=VAR_016336.
 /FTID=VAR_016337.
 /FTID=VAR_016338.
 /FTID=VAR_016339.
 /FTID=VAR_016340.
 /FTID=VAR_016341.
 /FTID=VAR_016342.
 /FTID=VAR_016343.
 /FTID=VAR_016344.
 /FTID=VAR_016345.
 /FTID=VAR_016346.
 /FTID=VAR_016347.
 /FTID=VAR_016348.
 /FTID=VAR_016349.
 /FTID=VAR_016350.
 /FTID=VAR_016351.
 /FTID=VAR_016352.
 /FTID=VAR_016353.
 /FTID=VAR_016354.
 /FTID=VAR_016355.
 /FTID=VAR_016356.
 /FTID=VAR_016357.
 /FTID=VAR_016358.
 /FTID=VAR_016359.
 /FTID=VAR_016360.
 /FTID=VAR_016361.
 /FTID=VAR_016362.
 /FTID=VAR_016363.
 /FTID=VAR_016364.
 /FTID=VAR_016365.
 /FTID=VAR_016366.
 /FTID=VAR_016367.
 /FTID=VAR_016368.
 /FTID=VAR_016369.
 /FTID=VAR_016370.
 /FTID=VAR_016371.
 /FTID=VAR_016372.
 /FTID=VAR_016373.
 /FTID=VAR_016374.
 /FTID=VAR_016375.
 /FTID=VAR_016376.
 /FTID=VAR_016377.
 /FTID=VAR_016378.
 /FTID=VAR_016379.
 /FTID=VAR_016380.
 /FTID=VAR_016381.
 /FTID=VAR_016382.
 /FTID=VAR_016383.
 /FTID=VAR_016384.
 /FTID=VAR_016385.
 /FTID=VAR_016386.
 /FTID=VAR_016387.
 /FTID=VAR_016388.
 /FTID=VAR_016389.
 /FTID=VAR_016390.
 /FTID=VAR_016391.
 /FTID=VAR_016392.
 /FTID=VAR_016393.
 /FTID=VAR_016394.
 /FTID=VAR_016395.
 /FTID=VAR_016396.
 /FTID=VAR_016397.
 /FTID=VAR_016398.
 /FTID=VAR_016399.
 /FTID=VAR_016400.
 /FTID=VAR_016401.
 /FTID=VAR_016402.
 /FTID=VAR_016403.
 /FTID=VAR_016404.
 /FTID=VAR_016405.
 /FTID=VAR_016406.
 /FTID=VAR_016407.
 /FTID=VAR_016408.
 /FTID=VAR_016409.
 /FTID=VAR_016410.
 /FTID=VAR_016411.
 /FTID=VAR_016412.
 /FTID=VAR_016413.
 /FTID=VAR_016414.
 /FTID=VAR_016415.
 /FTID=VAR_016416.
 /FTID=VAR_016417.
 /FTID=VAR_016418.
 /FTID=VAR_016419.
 /FTID=VAR_016420.
 /FTID=VAR_016421.
 /FTID=VAR_016422.
 /FTID=VAR_016423.
 /FTID=VAR_016424.
 /FTID=VAR_016425.
 /FTID=VAR_016426.
 /FTID=VAR_016427.
 /FTID=VAR_016428.
 /FTID=VAR_016429.
 /FTID=VAR_016430.
 /FTID=VAR_016431.
 /FTID=VAR_016432.
 /FTID=VAR_016433.
 /FTID=VAR_016434.
 /FTID=VAR_016435.
 /FTID=VAR_016436.
 /FTID=VAR_016437.
 /FTID=VAR_016438.
 /FTID=VAR_016439.
 /FTID=VAR_016440.
 /FTID=VAR_016441.
 /FTID=VAR_016442.
 /FTID=VAR_016443.
 /FTID=VAR_016444.
 /FTID=VAR_016445.
 /FTID=VAR_016446.
 /FTID=VAR_016447.
 /FTID=VAR_016448.
 /FTID=VAR_016449.
 /FTID=VAR_016450.
 /FTID=VAR_016451.
 /FTID=VAR_016452.
 /FTID=VAR_016453.
 /FTID=VAR_016454.
 /FTID=VAR_016455.
 /FTID=VAR_016456.
 /FTID=VAR_016457.
 /FTID=VAR_016458.
 /FTID=VAR_016459.
 /FTID=VAR_016460.
 /FTID=VAR_016461.
 /FTID=VAR_016462.
 /FTID=VAR_016463.
 /FTID=VAR_016464.
 /FTID=VAR_016465.
 /FTID=VAR_016466.
 /FTID=VAR_016467.
 /FTID=VAR_016468.
 /FTID=VAR_016469.
 /FTID=VAR_016470.
 /FTID=VAR_016471.
 /FTID=VAR_016472.
 /FTID=VAR_016473.
 /FTID=VAR_016474.
 /FTID=VAR_016475.
 /FTID=VAR_016476.
 /FTID=VAR_016477.
 /FTID=VAR_016478.
 /FTID=VAR_016479.
 /FTID=VAR_016480.
 /FTID=VAR_016481.
 /FTID=VAR_016482.
 /FTID=VAR_016483.
 /FTID=VAR_016484.
 /FTID=VAR_016485.
 /FTID=VAR_016486.
 /FTID=VAR_016487.
 /FTID=VAR_016488.
 /FTID=VAR_016489.
 /FTID=VAR_016490.
 /FTID=VAR_016491.
 /FTID=VAR_016492.
 /FTID=VAR_016493.
 /FTID=VAR_016494.
 /FTID=VAR_016495.
 /FTID=VAR_016496.
 /FTID=VAR_016497.
 /FTID=VAR_016498.
 /FTID=VAR_016499.
 /FTID=VAR_016500.
 /FTID=VAR_016501.
 /FTID=VAR_016502.
 /FTID=VAR_016503.
 /FTID=VAR_016504.
 /FTID=VAR_016505.
 /FTID=VAR_016506.
 /FTID=VAR_016507.
 /FTID=VAR_016508.
 /FTID=VAR_016509.
 /FTID=VAR_016510.
 /FTID=VAR_016511.
 /FTID=VAR_016512.
 /FTID=VAR_016513.
 /FTID=VAR_016514.
 /FTID=VAR_016515.
 /FTID=VAR_016516.
 /FTID=VAR_016517.
 /FTID=VAR_016518.
 /FTID=VAR_016519.
 /FTID=VAR_016520.
 /FTID=VAR_016521.
 /FTID=VAR_016522.
 /FTID=VAR_016523.
 /FTID=VAR_016524.
 /FTID=VAR_016525.
 /FTID=VAR_016526.
 /FTID=VAR_016527.
 /FTID=VAR_016528.
 /FTID=VAR_016529.
 /FTID=VAR_016530.
 /FTID=VAR_016531.
 /FTID=VAR_016532.
 /FTID=VAR_016533.
 /FTID=VAR_016534.
 /FTID=VAR_016535.
 /FTID=VAR_016536.
 /FTID=VAR_016537.
 /FTID=VAR_016538.
 /FTID=VAR_016539.
 /FTID=VAR_016540.
 /FTID=VAR_016541.
 /FTID=VAR_016542.
 /FTID=VAR_016543.
 /FTID=VAR_016544.
 /FTID=VAR_016545.
 /FTID=VAR_016546.
 /FTID=VAR_016547.
 /FTID=VAR_016548.
 /FTID=VAR_016549.
 /FTID=VAR_016550.
 /FTID=VAR_016551.
 /FTID=VAR_016552.
 /FTID=VAR_016553.
 /FTID=VAR_016554.
 /FTID=VAR_016555.
 /FTID=VAR_016556.
 /FTID=VAR_016557.
 /FTID=VAR_016558.
 /FTID=VAR_016559.
 /FTID=VAR_016560.
 /FTID=VAR_016561.
 /FTID=VAR_016562.
 /FTID=VAR_016563.
 /FTID=VAR_016564.
 /FTID=VAR_016565.
 /FTID=VAR_016566.
 /FTID=VAR_016567.
 /FTID=VAR_016568.
 /FTID=VAR_016569.
 /FTID=VAR_016570.
 /FTID=VAR_016571.
 /FTID=VAR_016572.
 /FTID=VAR_016573.
 /FTID=VAR_016574.
 /FTID=VAR_016575.
 /FTID=VAR_016576.
 /FTID=VAR_016577.
 /FTID=VAR_016578.
 /FTID=VAR_016579.
 /FTID=VAR_016580.
 /FTID=VAR_016581.
 /FTID=VAR_016582.
 /FTID=VAR_016583.
 /FTID=VAR_016584.
 /FTID=VAR_016585.
 /FTID=VAR_016586.
 /FTID=VAR_016587.
 /FTID=VAR_016588.
 /FTID=VAR_016589.
 /FTID=VAR_016590.
 /FTID=VAR_016591.
 /FTID=VAR_016592.
 /FTID=VAR_016593.
 /FTID=VAR_016594.
 /FTID=VAR_016595.
 /FTID=VAR_016596.
 /FTID=VAR_016597.
 /FTID=VAR_016598.
 /FTID=VAR_016599.
 /FTID=VAR_016600.
 /FTID=VAR_016601.
 /FTID=VAR_016602.
 /FTID=VAR_016603.
 /FTID=VAR_016604.
 /FTID=VAR_016605.
 /FTID=VAR_016606.
 /FTID=VAR_016607.
 /FTID=VAR_016608.
 /FTID=VAR_016609.
 /FTID=VAR_016610.
 /FTID=VAR_016611.
 /FTID=VAR_016612.
 /FTID=VAR_016613.
 /FTID=VAR_016614.
 /FTID=VAR_016615.
 /FTID=VAR_016616.
 /FTID=VAR_016617.
 /FTID=VAR_016618.
 /FTID=VAR_016619.
 /FTID=VAR_016620.
 /FTID=VAR_016621.
 /FTID=VAR_016622.
 /FTID=VAR_016623.
 /FTID=VAR_016624.
 /FTID=VAR_016625.
 /FTID=VAR_016626.
 /FTID=VAR_016627.
 /FTID=VAR_016628.
 /FTID=VAR_016629.
 /FTID=VAR_016630.
 /FTID=VAR_016631.
 /FTID=VAR_016632.
 /FTID=VAR_016633.
 /FTID=VAR_016634.
 /FTID=VAR_016635.
 /FTID=VAR_016636.
 /FTID=VAR_016637.
 /FTID=VAR_016638.
 /FTID=VAR_016639.
 /FTID=VAR_016640.
 /FTID=VAR_016641.
 /FTID=VAR_016642.
 /FTID=VAR_016643.
 /FTID=VAR_016644.
 /FTID=VAR_016645.
 /FTID=VAR_016646.
 /FTID=VAR_016647.
 /FTID=VAR_016648.
 /FTID=VAR_016649.
 /FTID=VAR_016650.
 /FTID=VAR_016651.
 /FTID=VAR_016652.
 /FTID=VAR_016653.
 /FTID=VAR_016654.
 /FTID=VAR_016655.
 /FTID=VAR_016656.
 /FTID=VAR_016657.
 /FTID=VAR_016658.
 /FTID=VAR_016659.
 /FTID=VAR_016660.
 /FTID=VAR_016661.
 /FTID=VAR_016662.
 /FTID=VAR_016663.
 /FTID=VAR_016664.
 /FTID=VAR_016665.
 /FTID=VAR_016666.
 /FTID=VAR_016667.
 /FTID=VAR_016668.
 /FTID=VAR_016669.
 /FTID=VAR_016670.
 /FTID=VAR_016671.
 /FTID=VAR_016672.
 /FTID=VAR_016673.
 /FTID=VAR_016674.
 /FTID=VAR_016675.
 /FTID=VAR_016676.
 /FTID=VAR_016677.
 /FTID=VAR_016678.
 /FTID=VAR_016679.
 /FTID=VAR_016680.
 /FTID=VAR_016681.
 /FTID=VAR_016682.
 /FTID=VAR_016683.
 /FTID=VAR_016684.
 /FTID=VAR_016685.
 /FTID=VAR_016686.
 /FTID=VAR_016687.
 /FTID=VAR_016688.
 /FTID=VAR_016689.
 /FTID=VAR_016690.
 /FTID=VAR_016691.
 /FTID=VAR_016692.
 /FTID=VAR_016693.
 /FTID=VAR_016694.
 /FTID=VAR_016695.
 /FTID=VAR_016696.
 /FTID=VAR_016697.
 /FTID=VAR_016698.
 /FTID=VAR_016699.
 /FTID=VAR_016700.
 /FTID=VAR_016701.
 /FTID=VAR_016702.
 /FTID=VAR_016703.
 /FTID=VAR_016704.
 /FTID=VAR_016705.
 /FTID=VAR_016706.
 /FTID=VAR_016707.
 /FTID=VAR_016708.
 /FTID=VAR_016709.
 /FTID=VAR_016710.
 /FTID=VAR_016711.
 /FTID=VAR_016712.
 /FTID=VAR_016713.
 /FTID=VAR_016714.
 /FTID=VAR_016715.
 /FTID=VAR_016716.
 /FTID=VAR_016717.
 /FTID=VAR_016718.
 /FTID=VAR_016719.
 /FTID=VAR_016720.
 /FTID=VAR_016721.
 /FTID=VAR_016722.
 /FTID=VAR_016723.
 /FTID=VAR_016724.
 /FTID=VAR_016725.
 /FTID=VAR_016726.
 /FTID=VAR_016727.
 /FTID=VAR_016728.
 /FTID=VAR_016729.
 /FTID=VAR_016730.
 /FTID=VAR_016731.
 /FTID=VAR_016732.
 /FTID=VAR_016733.
 /FTID=VAR_016734.
 /FTID=VAR_016735.
 /FTID=VAR_016736.
 /FTID=VAR_016737.
 /FTID=VAR_016738.
 /FTID=VAR_016739.
 /FTID=VAR_016740.
 /FTID=VAR_016741.
 /FTID=VAR_016742.
 /FTID=VAR_016743.
 /FTID=VAR_016744.
 /FTID=VAR_016745.
 /FTID=VAR_016746.
 /FTID=VAR_016747.
 /FTID=VAR_016748.
 /FTID=VAR_016749.
 /FTID=VAR_016750.
 /FTID=VAR_016751.
 /FTID=VAR_016752.
 /FTID=VAR_016753.
 /FTID=VAR_016754.
 /FTID=VAR_016755.
 /FTID=VAR_016756.
 /FTID=VAR_016757.
 /FTID=VAR_016758.
 /FTID=VAR_016759.
 /FTID=VAR_016760.
 /FTID=VAR_016761.
 /FTID=VAR_016762.
 /FTID=VAR_016763.
 /FTID=VAR_016764.
 /FTID=VAR_016765.
 /FTID=VAR_016766.
 /FTID=VAR_016767.
 /FTID=VAR_016768.
 /FTID=VAR_016769.
 /FTID=VAR_016770.
 /FTID=VAR_016771.
 /FTID=VAR_016772.
 /FTID=VAR_016773.
 /FTID=VAR_016774.
 /FTID=VAR_016775.
 /FTID=VAR_016776.
 /FTID=VAR_016777.
 /FTID=VAR_016778.
 /FTID=VAR_016779.
 /FTID=VAR_016780.
 /FTID=VAR_016781.
 /FTID=VAR_016782.
 /FTID=VAR_016783.
 /FTID=VAR_016784.
 /FTID=VAR_016785.
 /FTID=VAR_016786.
 /FTID=VAR_016787.
 /FTID=VAR_016788.
 /FTID=VAR_016789.
 /FTID=VAR_016790.
 /FTID=VAR_016791.
 /FTID=VAR_016792.
 /FTID=VAR_016793.
 /FTID=VAR_016794.
 /FTID=VAR_016795.
 /FTID=VAR_016796.
 /FTID=VAR_016797.
 /FTID=VAR_016798.
 /FTID=VAR_016799.
 /FTID=VAR_016800.
 /FTID=VAR_016801.
 /FTID=VAR_016802.
 /FTID=VAR_016803.
 /FTID=VAR_016804.
 /FTID=VAR_016805.
 /FTID=VAR_016806.
 /FTID=VAR_016807.
 /FTID=VAR_016808.
 /FTID=VAR_016809.
 /FTID=VAR_016810.
 /FTID=VAR_016811.
 /FTID=VAR_016812.
 /FTID=VAR_016813.
 /FTID=VAR_016814.
 /FTID=VAR_016815.
 /FTID=VAR_016816.
 /FTID=VAR_016817.
 /FTID=VAR_016818.
 /FTID=VAR_016819.
 /FTID=VAR_016820.
 /FTID=VAR_016821.
 /FTID=VAR_016822.
 /FTID=VAR_016823.
 /FTID=VAR_016824.
 /FTID=VAR_016825.
 /FTID=VAR_016826.<

Db	218	RULT--PDAFVCR-----TICTYHFN--CLFFTFYT-----NVWKIESQRNV	255
Qy	112	---KCSNSETECDS-----SG-----TCINP-----SNWCDGVSHCPGEGDENRCVRL	151
Db	256	CLLKTBESGTFSSSTPOENTISYSLTCKRTLPBCHSKIYPGVDF--GGEELN-----	308
Qy	152	YGFNFILQVYVSORKSMWHPVCDDWENYGRAACRDMGYKNNFYSSCGIIVDDSGSTS----	208
Db	309	--VTFVKGV-----NVCQE-----TCTKM--IRCQFTYSLLPEDCKEKKKC	347
Qy	209	FMKLNTSAGNVDIYKLYHSDACSSKAVVSLRCIACGVN--LNSSRQSRIVGSESALPGA	266
Db	348	FLRLMDGSPTRL---AYGTQSSG---YSLRLCNTGDNVCTKTSTRIVGTNSSWGE	401
Qy	267	WPQVSLHVQ---NVHVCGSIIITPEWIVTAACHVE--KPLNNPWHWTAFAGILRQSFMY	322
Db	402	WPQVSLQVKLTAQRHLCGSLIGHQWVLTAAHCFDGLPLQDVWR--IYSGILNLSDITK	459
Qy	323	GAGY-QVEKVISHENYDSKTNNDIALMKLOKPLTFNDLVKPYCLNPNPQWMLQPEQLCWI	381
Db	460	DTPFSQIKETIIHQNYKVSEGNHDIALIKLOAPLNYTEFQKPICLPSKGTSTIYTNCWV	519
Qy	382	SGWATEEKGKTSBVLNAAKVLLIETQCNRSYVYDNLITPAMICAGFLQGNVDSQCGDS	441
Db	520	TGWGFSKEKEIQNILQWNIPLVTNEECQKRY-QDYKITQRMVCAGYKEGKDKACKGDS	578
Qy	442	GGPLVTSKNNIWWLIGTSGSGCAKAYRPGVYGNVWVFTDWIYRQMR-ADG	492
Db	579	GGPLVCKHNGMWRLVGITSWGEGCARREQPGVYTKVAEYNDWILEKTQSSDG	630

Search completed: June 1, 2004, 14:35:07
Job time : 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 14:20:43 ; Search time 45 Seconds
(without alignments)
1051.694 Million cell updates/sec

Title: US-09-615-285B-2

Perfect score: 2717

Sequence: 1 MALNSGSPPAIGPVYENHGY.....VYGVNWVFTDIYRQMRADG 492

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 78:**

2: PIR1:**

3: PIR2:**

4: PIR3:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	660	24.3	1035	1 A43090	enteropeptidase (E
2	647	23.8	1019	1 A56318	enteropeptidase (E
3	634	23.3	1034	1 A53663	enteropeptidase (E
4	577.5	21.3	1113	2 J6315	low-density lipopr
5	569.5	21.0	638	1 KQSDP	plasma kallikrein
6	565.5	20.8	417	1 S09845	hepsin (EC 3.4.21.
7	564.5	20.8	855	2 J0731	membrane-bound arg
8	556	20.5	638	1 KQRTPL	plasma kallikrein
9	538	19.8	638	1 KQMSPL	plasma kallikrein
10	533	19.6	625	1 KPDU1	coagulation factor
11	529.5	19.5	416	1 S3377	hepsin (EC 3.4.21.
12	518.5	19.1	812	1 PLMS	plasmin (EC 3.4.21
13	501	18.4	421	1 S11674	acrosin (EC 3.4.21
14	501	18.4	1524	2 T30337	polyprotein - Afri
15	490.5	18.1	415	1 A34170	acrosin (EC 3.4.21
16	489.5	18.0	436	2 JX0172	acrosin (EC 3.4.21
17	486.5	17.9	761	2 JCS759	brain-specific ser
18	483	17.8	418	2 A37344	acrosin (EC 3.4.21
19	477	17.6	790	1 PLG3	plasmin (EC 3.4.21
20	475.5	17.5	431	2 S47536	acrosin (EC 3.4.21
21	473.5	17.4	421	2 S29599	acrosin (EC 3.4.21
22	472.5	17.4	810	2 B30848	plasmin (EC 3.4.21
23	470.5	17.3	437	2 S18407	acrosin (EC 3.4.21
24	470.5	17.3	810	2 I46260	plasmin (EC 3.4.21
25	461	17.0	343	2 A57014	proctasin (EC 3.4.
26	456.5	16.8	460	2 B61545	plasmin (EC 3.4.21
27	455	16.7	812	1 PLSO	plasmin (EC 3.4.21
28	451	16.6	2616	2 S47096	nudel protein prec
29	450.5	16.6	267	2 S40006	trypsin (EC 3.4.21

RESULT 1

A43090

enteropeptidase (EC 3.4.21.9) precursor [validated] - bovine

N;Alternate names: enterokinase

C;Species: Bos primigenius taurus (cattle)

C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 28-Apr-2003

C;Accession: A43090; A48874; A61436

R;Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.

Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994

A;Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease comp

A;Reference number: A43090; MUID:94329561; PMID:8052624

A;Accession: A43090

A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1035 <KIT>

A;Cross-references: GB:U09859; NID:9746410; PIDN:ARB40026.1; PID:9746411

A;Experimental source: small intestine

R;Avallie, E.R.; Renemulla, A.; Racie, L.A.; DiBlasio, E.A.; Perez, C.; Grant, K.L.;

J. Biol. Chem. 269, 23311-23317, 1993

A;Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of I

A;Reference number: A48874; MUID:94043122; PMID:8226855

A;Accession: A48874

A;Molecule type: mRNA

A;Residues: 801-1035 <LAV>

A;Cross-references: GB:L19663; NID:9416131; PIDN:AAA16035.1; PID:9416132

A;Note: parts of this sequence, including the amino end of the mature protein, were conl

R;Light, A.; Janska, H.

J. Protein Chem. 10, 475-480, 1991

A;Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.

A;Reference number: A61436; MUID:92189715; PMID:1799406

A;Accession: A61436

A;Molecule type: protein

A;Residues: 801-807, 'Y', 809-827 <LIG>

C;Comment: The mechanism of association with the membrane of the intestinal brush border

membrane attachment using a signal-anchor sequence.

C;Comment: Conversion from membrane-bound to soluble forms may involve further processin

C;Complex: mature enteropeptidase is variously reported to contain two (heavy and light)

lptide linked

C;Function:

A;Description: cleaves propeptide from trypsinogen to produce active trypsin

A;Pathway: intestinal digestive hydrolase cascade

C;Superfamily: enteropeptidase; C1x/C1s repeat homology; LDL receptor ligand-binding ref

C;Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protei

F;22-38/Domain: transmembrane #status predicted <TM>

F;52-117/Product: enteropeptidase mini chain #status predicted <MCH>

F;118-800/Product: enteropeptidase heavy chain #status predicted <HCH>

F;199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F;358-520/Domain: MAM homology <MAM>

F;542-647/Domain: C1x/C1s repeat homology <C1R>

F;659-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F;844-799/Domain: scavenger receptor cysteine-rich domain homology <SRC

F;801-1035/Product: enteropeptidase light chain #status predicted <LCH>

mast cell trypsinase
trypsin (EC 3.4.21
hepatocyte growth
trypsin (EC 3.4.21
mast cell proteinase
acrosin (EC 3.4.21
plasmin (EC 3.4.21
plasmin (EC 3.4.21
apoptein(a) (EC
trypsinase (EC 3.4.2
serine proteinase
trypsinase (EC 3.4.2
trypsin (EC 3.4.21
trypsin (EC 3.4.21
trypsin (EC 3.4.21

ALIGNMENTS

30	450	16.6	270	2	S56160
31	449	16.5	275	2	S40005
32	449	16.5	655	1	A46688
33	448	16.5	277	2	S35340
34	447	16.5	276	2	A38654
35	446.5	16.4	420	2	A55283
36	445.5	16.4	810	1	PLHU
37	444.5	16.4	455	2	A61545
38	443.5	16.3	266	2	S54146
39	442.5	16.3	458	1	S00657
40	441.5	16.2	274	2	JC4171
41	441	16.2	786	1	A47547
42	439	16.2	273	2	A47246
43	438	16.1	274	2	S35339
44	437	16.1	275	2	S40007
45	436	16.0	242	2	S49489

F:801-1030/Domain: trypsin homology <TRY>
F:115,147,170,194,233,263,264,456,486,519,550,646,698,722,741,762,864,903,965/Binding
F:788-912,826-842,926-933,957-972,983-1011/Disulfide bonds: #status predicted
F:841,892,987/Active site: His, Asp, Ser #status predicted

Query Match 24.3%; Score 660; DB 1; Length 1035;
Best Local Similarity 33.9%; Pred. No. 8e-40;
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;

QY 65 VCTQKSPSTVCTSKTKALCTITL-----GTELVGAALAGLWKGSKCSNSG 117
DB 611 VITGPGVNVDFSTTRMTVLFITDNMLAKQFKANFTTYGLG-----IPPECKEDN 663
QY 118 IECDSSGTCINPNSWCDGVSHCPGGEDENRCVRLYG-----PNFLOVYSSQKSWH 169
DB 664 FQC-KDGEICPLVNLCDGPFCHCKDGSDEARCVLRFNGITDSSGLVQVRIQ-----SIWH 716
QY 170 PVQDDNNENYGRAACDMGYKNFYSSQGIYDDSGSTSFMKLNTSAGNVDIYKLYHSD 229
DB 717 VACAENWTTQISDDVQCLGLGTG---NSSVPFTSGGPGYVNLNTAPNGSLI---LPSQ 771
QY 230 ACSKSAVVSURC--IACGVNLSSROS-RIVGGESALPGAMPQVSLHVQNVHYCGSII 286
DB 772 QCLEDSLILLCQNYKSCGKKLVQEVSPKIVGSDSRREGAMPVVALYFDDQVCGASLV 831
QY 287 TSPWIVTAACHVEKPLNPNHMTAFAGILRQSFMYGAGYVE-----KVISHENYDSKT 341
DB 832 SRDLVSAACHVYGRNWSKKAVALGLHNASNL---TSPQIETRLIDQIVINPHYNKR 888
QY 342 KNNDIALMKLQKPLTENDLVKVPCLPNPGMQLQPLQCLWISGWGATEKGTSEVLA 401
DB 889 KNNDIAMHLEMKVNTDYIQICLPEENQVPPGRCISAGGALIVQGSTADVLQ 948
QY 402 VLLIETORCNSRVYVNDLITPAMICAGFLQGVNDVSCQDGGPLVTSKNNTWLLIGT 461
DB 949 VPLISNEKCOQQQPEYN-ITENMVCAGYEAGGVDSQDGGPLVTSKNNTWLLIGT 1007
QY 462 GSGCAKAYRPGVYVNVVFTDWI 484
DB 1008 GYQCALPNRPGVYVNVVFTDWI 1030

RESULT 2
A56318
enteropeptidase (EC 3.4.21.9) precursor [validated] - human
N/Alternate names: enterokinase
C/Species: Homo sapiens (man)
C/Date: 19-May-1995 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003
C/Accession: A56318; B43090
R/Kitamoto, Y.; Veille, R.A.; 1995
Biochemistry 34, 4562-4568, 1995
A/Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolytic
A/Reference number: A56318; MUID:95234679; PMID:7718557
A/Accession: A56318
A/Molecule type: mRNA
A/Residues: 1-1019 <KIT>
A/Cross-references: GB:U09860; NID:9746412; PIDN:AA050138.1; PID:g746413
R/Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
A/Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
A/Reference number: A43090; MUID:94329561; PMID:8052624
A/Accession: B43090
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 749-1019 <K12>
A/Cross-references: GB:U09860
C/Comment: The mechanism of association with the membrane of the intestinal brush border
ctated below) or with amino-terminal myristoylation of the heavy chain.
A/Gene: GDB:PRSS7
A/Cross-references: GDB:384083; OMIM:226200
A/Map position: 21q21-21q21
C/Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)

ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms invol
ducts.
C/Function:
A/Description: cleaves activation peptide from trypsinogen to produce active trypsin
A/Pathway: intestinal digestive hydrolase cascade
C/Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding r
C/Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F:1-784/Product: enteropeptidase heavy chain #status predicted <HC>
F:22-38/Domain: transmembrane #status predicted <TM>
F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:342-504/Domain: NAM homology <NAM>
F:526-531/Domain: C1r/C1s repeat homology <C1r>
F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SR
F:785-1019/Product: enteropeptidase light chain #status predicted <LC>
F:785-1014/Domain: trypsin homology <TRY>
F:116,147,179,328,335,388,440,470,503,630,682,706,725,848,887,909,949/Binding site:
F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted
F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 23.8%; Score 647; DB 1; Length 1019;
Best Local Similarity 35.8%; Pred. No. 6.8e-39;
Matches 138; Conservative 67; Mismatches 154; Indels 26; Gaps 10;

QY 113 CNSGIECDSSGTCINPNSWCDGVSHCPGGEDENRCVRLYG-----ENFILQVYSSQ 164
DB 643 CXADHFQC-KNGECVPLVNLCDGLHCKEDGSDADCVRFNGTNNGLVRFRIQ----- 696
QY 165 RKSHPVQDDNNENYGRAACDMGYKNFYSSQGIYDDSGSTSFMKLNTSAGNVDIYK 224
DB 697 -SIWHTACAEANWTTQISNDVQCLGLGSG-NSSKPIFTSDGG-PFVKLNTAP---DGHLLI 750
QY 225 LVHSDACSSKAVSVLR--IACGVNLSSR-QSRIVGGESALPGAMPQVSLHVQNVHVC 281
DB 751 LPSQCLQDLSLRLQCNHKSCKGLAAQDITPKLVGSKNAKEGAMPVWGLYGGELLC 810
QY 282 GSGIITPWIIVTAACHVEKPLNPNHMTAFAGILRQSFMP--YGAGVQVEKVISHPNYDS 339
DB 811 GASLVSSDWLYSAACHVYGRNLEPSKWTAILGLHMKNLATSPQVPLDIIEVINPHYNR 870
QY 340 KTKNDIALMKLQKPLTENDLVKVPCLPNPGMQLQPLQCLWISGWGATEKGTSEVLA 399
DB 871 RKNDIAMHLEKPVNTDYIQICLPEENQVPPGRCISAGGALIVQGSTADVLQ 930
QY 400 AKVLLIETORCNSRVYVNDLITPAMICAGFLQGVNDVSCQDGGPLVTSKNNTWLLIGDT 459
DB 931 ADVPLLSNERCQQQMPEN-ITENMICAGYEGGIDSCQDGGPLVTSKNNTWLLIGDT 989
QY 460 SMGSGCAKAYRPGVYVNVVFTDWI 484
DB 990 SFGYKCALPNRPGVYVNVVFTDWI 1014

RESULT 3
A53663
enteropeptidase (EC 3.4.21.9) precursor [validated] - pig
N/Alternate names: enterokinase
C/Species: Sus scrofa domestica (domestic pig)
C/Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003
C/Accession: A53663
R/Watsushima, M.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada, S.; Miki, K.; Kurokawa,
J. Biol. Chem. 269, 19976-19982, 1994
A/Title: Structural characterization of porcine enteropeptidase.
A/Reference number: A53663; MUID:94327548; PMID:8051081
A/Accession: A53663
A/Molecule type: mRNA
A/Residues: 1-1034 <MAT>
A/Cross-references: GB:D30799; NID:9505122; PIDN:BA06459.1; PID:g505123
A/Note: parts of this sequence, including the amino ends of three chains isolated from
C/Comment: The mechanism of association with the membrane of the intestinal brush borde
ctated below) or with amino-terminal myristoylation of the heavy chain.
C/Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms invol

C;Function: cleaves activation peptide from trypsinogen to produce active trypsin
 A;Pathway: intestinal digestive hydrolase cascade
 C;Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding repeat
 C;Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
 F;22-38/Domain: transmembrane #status predicted <TM>
 F;52-117/Product: enteropeptidase mini chain #status predicted <MCH>
 F;118-799/Product: enteropeptidase heavy chain #status predicted <HCH>
 F;119-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F;357-519/Domain: MAM homology <MAM>
 F;541-646/Domain: C1r/C1s repeat homology <C1R>
 F;658-692/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F;693-798/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC>
 F;800-1034/Product: enteropeptidase light chain #status predicted <LCH>
 F;800-1029/Domain: trypsin homology <TRY>
 F;116-147/Domain: trypsin homology <TRY>
 F;116-147/170-194-283-343-350-403-455-518-549-645-697-701-721-740-761-804-863-902-96
 F;787-911-825-841-925-982-986-971-982-1010/Disulfide bonds: #status predicted
 F;840-891-986/Active site: His, Asp, Ser #status predicted

Query Match 23.3%; Score 634; DB 1; Length 1034;
 Best Local Similarity 33.3%; Pred. No. 6.1e-38;
 Matches 147; Conservative 72; Mismatches 179; Indels 44; Gaps 13;
 QY 65 VCTQKSPSGTCTKTKALCIT---ITLG-----TFLVGAALAGLLKMGKSCNSG 117
 DB 610 VYTGPGVEDVFTTNRMTVLTINDALTKGFKANFTTGYHLG-----IPEPCKEDN 662
 QY 118 IECDSSGTCINPSNWCDSVSHCPGSGDENRCVRLYG--PNFTIQVYSSQKSWHPVCDD 175
 DB 663 FQCE-NGECVLLVNLCDGFSCHKDGSDEAHCVRLNGTANNGLVQFRIQSIWHTACAA 721
 QY 176 WNEYGEACRDGMKYN-----FYSQGIIVDSGSTSPMKLNTSAGNVDIYKLIYHSDA 230
 DB 722 WTTQTSDVDCQLGLGTGNSMPFFS-----SGGGPFVKLNTAPNGSLI---LTASEQ 771
 QY 231 CSSKAVVSLRC--IAGCVNLNSRQS-RIVGGESALPGAMPQVSLHVNQVHVCSSIT 287
 DB 772 CFEDSLILLCNHKSCGKQVAGVSPKIVGNDREGAPVWVALYNGQLLOGASIVS 831
 QY 288 PEMIVTAACHVEKPLNPHWTAFAGILRQSPFMFVAGVQ-----VEKVIHPNVDSTKY 342
 DB 832 RDVLVSAACHVYGRNLEPSKVAIGLHMTSL---TSPQIVTRLIDIDINPHNRRK 888
 QY 343 NNDIALMKLQKPLTFNDLVKPLCLPNPMMLOPEQLCWISGATGATEEKGKTSSEVLNAKV 402
 DB 889 DSDIAMHLEFKVNYTDIOPICLPEENQVFPFGRICTSIAGKVIYQSGSPADILQEA 948
 QY 403 LLIEIQRCSRYVDNLITPAMICAGFLQGNVDSQSGSGGLPVTSKNNIWLIGDTSWG 462
 DB 949 PLUSNEKCCQQMPEYN-ITENMWCAGYEGGIDSCQSGSGGLPMLCLENRRLLAGVTSFG 1007
 QY 463 SGCAKAYRPGVYGNVMVFTDWI 484
 DB 1008 YQCALPNREGVYARVPKFTWI 1029

RESULT 4
 JB0315
 low-density lipoprotein receptor-related protein - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Feb-2003
 C;Accession: JE0315
 F;Tomita, Y.; Kim, D.; Magoori, K.; Fujino, T.; Yamamoto, T.T.
 J. Biochem. 124, 784-789, 1998
 A;Title: A novel low-density lipoprotein receptor-related protein with type II membrane
 A;Reference number: JE0315; MUID:98429596; PMID:9756624
 A;Accession: JE0315
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-1113 <TM>
 A;Cross-references: DBJ:AB013874; NID:93869144; PIDN:BA034371.1; PID:93869145
 C;Superfamily: mouse low-density lipoprotein receptor-related protein; LDL receptor liga
 F;337-372/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F;374-408/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F;410-445/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F;447-482/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F;648-682/Domain: LDL receptor ligand-binding repeat homology <LDL5>
 F;684-720/Domain: LDL receptor ligand-binding repeat homology <LDL6>
 F;723-757/Domain: LDL receptor ligand-binding repeat homology <LDL7>
 F;869-1097/Domain: trypsin homology <TRY>

Query Match 21.3%; Score 577.5; DB 2; Length 1113;
 Best Local Similarity 33.3%; Pred. No. 8.1e-34;
 Matches 134; Conservative 67; Mismatches 155; Indels 47; Gaps 14;

QY 111 SKCSNGIECDSSGTCINPSNWCDSVSHCPGSGDENRCVRLY---GNFTIQVYSSQK 167
 DB 721 SFQDNELAC-ANHECVPEDLACDGVDCSDSSDMGCVTLKNGNSSLLTVHKSAXE- 778
 QY 168 WHPVCDDNWNHYGRACRDGMKYNFYSQGIIVDSGSTSPMKLNTSAGNVD---IYKK 224
 DB 779 -HRVCADGWRETLSQLACKQMGLEP--SVTKLIFGQEQQLWRLPYWNLNLTQEL 835
 QY 225 LVHSDACSSKAVVSLRCIA--CGVNLNSRQSRIVGGESALPGAMPQVSLHVC-NVHVC 281
 DB 836 LVYHSCPSERSISLLCSKQDCGRPPAARMKRIILGRTSRPGRWPQCISQSEPSGHIC 895
 QY 282 GSGIITPEWIVTAACHVE-----KPLNPHWTAFAGILRQSPFMFVAGVQ 327
 DB 896 GCYLIARAKWLVVAHCFEGREDADYKVVFGINNLDHP-----SGFMQTRF- 941
 QY 328 VEKVIHPNVDSTKKNIDIALMKLQKPLTFNDLVKPLCLPNPMMLOPEQLCWISGAT 387
 DB 942 VKTILLHPYSRAVDYDISVVELSDDINETSYPVCLPSPPEVLEPDTYCYITGNG-- 999
 QY 388 BEKGTSEVLNAKVLLIETQRCSRYVDNLITPAMICAGFLQGNVDSQSGSGGLPVT 447
 DB 1000 HMGKNKPFLLQGEVRIIPLEQCS-YFDMKTITNRMICAGVESGTVDSCMGDSGGLVC 1058
 QY 448 SK-NNIWLIGDTSWGSGC-AKAYRPGVYGNVMVFTDWIYRQM 488
 DB 1059 ERPGGQWTLFGLTSWGSVCFSKVLGPGVYNSVYFVGWIERQI 1101

RESULT 5

KOHUP

plasma kallikrein (EC 3.4.21.34) precursor - human
 N;Alternate names: kininogenin; plasma prekallikrein
 C;Species: Homo sapiens (man)
 C;Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999
 C;Accession: A00921; A37939
 R;Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.
 Biochemistry 25, 2410-2417, 1986
 A;Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four t
 A;Reference number: A00921; MUID:86243359; PMID:3521732
 A;Accession: A00921
 A;Molecule type: mRNA
 A;Residues: 1-638 <CHU>
 A;Cross-references: GB:M13143; NID:9190262; PIDN:AAA60153.1; PID:9190263
 R;McMullen, B.A.; Fujikawa, K.; Davie, E.W.
 Biochemistry 30, 2050-2056, 1991
 A;Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of
 A;Reference number: A37939; MUID:91152016; PMID:1998666
 A;Accession: A37939
 A;Molecule type: protein
 A;Residues: 20-2740-46; 'X' 48, 'H' 50, 'X' 52-70, 'H' 75-76, 'X' 78-80, 103-113, 131-140, 141-
 1260-483, 'X' 285, 287-291, 'X' 293-295, 314-317, 'X' 319-320, 321-324, 'X' 329-333, 334-339, 'X'
 525, 538-551, 562, 'X' 564-567, 573, 'X' 575-576, 578-583, 'X' 585, 592-604 <MC>
 C;Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
 C;Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a 11
 are linked by one or more disulfide bonds.
 C;Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal r
 inogen and may also play a role in the renin-angiotensin system by converting prorenin i
 C;Genetics:
 A;Gene: GDB:KLK3
 A;Cross-references: GDB:127575; OMIM:229000

A:Map position: 4q35-4q35

C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-638/Product: plasma kallikrein #status predicted <NAT>
F:20-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:231-380/Domain: apple repeat <AP4>
F:391-621/Domain: plasma kallikrein light chain #status predicted <LCH>
F:391-621/Domain: trypsin homology <TRY>
F:21-104,47-77,51-194,137-166,141-147,201-284,227-256,231-237,292-375,322-328,363
F:127,308,396,453,494/Binding site: carbohydrate (Asn) #status experimental
F:318-347,340-345/Disulfide bonds: #status predicted
F:390-391/Cleavage site: Arg-11e (coagulation factor XIa) #status predicted
F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 21.0%; Score 569.5; DB 1; Length 638;
Best Local Similarity 32.8%; Pred. No. 1.6e-33;
Matches 155; Conservative 64; Mismatches 160; Indels 93; Gaps 23;

QY 55 RVLTAQSNPVCTQKSPSGTCTSKKALCITLTGLTFLVGAALAAAGLLKWKMGWS---111
DB 218 RVLTAQSNPVCTQKSPSGTCTSKKALCITLTGLTFLVGAALAAAGLLKWKMGWS---111
QY 112 ---KCSNGIECDSS-----SG-----TCINP-----SNMCDGVSHCPGGEDENRCVRL 151
DB 256 CLKLTSEGTSPSTPQENTISGYSLTCKRTLPCHSKIPYGVDF--GGEELN-----308
QY 152 YGFNFILQVYSQKSWHPVCDQDNNYNGRAACDMGKNNFYSSQGIIVDSGSTS---208
DB 309 --VTFVKGV-----NVQDS-----TCYK-IRCOFFYSLLPDCKEERCKC 347
QY 209 FMKLTASAGNVDYKKLYSHDACSAAVSLRCIACGVN--LNSRQRIRIVGSSALPGA 266
DB 348 FLRLSDMGSPTRI---AYGTQSSG---YSLRLCNTGNSVCTTKTSTRIVGVGTSNNGE 401
QY 267 WPMQVSLHQV---NVHVCVGSIIITPEWIVTAACHVE-KPLNPNMHTAFAGILRQSFMY 322
DB 402 WPMQVSLHQVTAQRHLGSSLIHQWVLTAAHCDGLPLQDWR--IYSGILNLSLDTK 459
QY 323 GAGY-QUEKVI SHPNVDSKNDIALMKQLPLFNDLVKPVCLPNFGMMLQPEQLCWI 381
DB 460 DTPFSQIKETIIHQNTKVSSEGNHDIALLKQAPLNNTYEFQKPICLPSKGDSTTIYTCGW 519
QY 382 SGMGATEEKGKTSSEVLNAAKVLIIETQRCNSRYVDNLITPAMICAGFLQGNVDSQGD 441
DB 520 TGHGFSKKEGEIQNILOKYNIPLVNTEECQKEY-QDYKIQPMVCAGYKEGKCKACKGDS 578
QY 442 GGLVTSKNNIWLIGDTSWGSCKAKAYRPGYGVNMVFTDWYKQMR-ADG 492
DB 579 GGLVTSKNNIWLIGDTSWGSCKAKAYRPGYGVNMVFTDWYKQMR-ADG 492

RESULT 6

S00845
hepsin (EC 3.4.21.-) - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1998 #sequence_revision 31-Dec-1998 #text_change 18-Jun-1999
C:Accession: S00845
R:Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davies, E.W.
Biochemistry 27, 1067-1074, 1988
A:Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane dom
A:Reference number: S00845; MUID:88209431; PMID:2835076
A:Accession: S00845
A:Molecule type: mRNA
A:Residues: 1-417 <LEY>
A:Cross-references: EMBL:X07732; NID:g32063; PIDN:CAA30558.1; PID:g32064
C:Genetics:
A:Gene: GDB:HPN; TMAPRSS1; hepsin
A:Cross-references: GDB:135685; OMIM:142440
A:Map position: 19q11-19q13.2

C:Superfamily: hepsin; trypsin homology
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F:23-45/Domain: transmembrane #status predicted <TMN>
F:163-400/Domain: trypsin homology <TRY>
F:188-204,291-359,322-338,349-381/Disulfide bonds: #status predicted
F:203,257,353/Active site: His, Asp, Ser #status predicted

Query Match 20.8%; Score 565.5; DB 1; Length 417;
Best Local Similarity 30.7%; Pred. No. 1.9e-33;
Matches 133; Conservative 68; Mismatches 173; Indels 59; Gaps 9;

QY 77 CTSKTKKALCITLTGLTFLVGAALAAAGLLKWKMGSKCSNGIECDSSGTCINPNWCDGV 136
DB 12 CCSRPKVA---ALTAGTLLTLLTAIGA-----ASWAIVA 41
QY 137 SHCPGGEDENRCVRLYGVNFILQVYSQKSWHPVCDQDNNYNGRAACDMGKNNFYSS 196
DB 42 VLLRSDQEPFLPVQVSSADARLMVDFKTEGTRLLCSRSNARVAGLSCEMGFLRALTH 101
QY 197 SQGIIVDSGSTSFMKLTASAGNVDYK-KLYHSDA-----CSSKAVVSLRCIACGV 246
DB 102 SELDVRTAGA-----NGTSGFFCVDGRLPHTQRLLEVISVCDPRFLAAICQDGG- 154
QY 247 NLNSRQRIRIVGSSALPGAAMPQVSLHQVNVHVCVGSIIITPEWIVTAACHVEKPLNPNW 306
DB 155 -RRKLPVTRIVGRTSLGRHPWQVSLRYDGAHLGSSLSGDMVLTAAHCFPERNVLS 213
QY 307 HWTAFAGILRQSFMYGAGYQVEKVISHPNY-----DSKTPNNDIALMKQLKPLTNDL 360
DB 214 RMRVFPAGAVQASP-HGLQLGVQAVVYHGGYLPFRDPNPNSEENSNDIALVHLSPLPTEY 272
QY 361 VPKVCLPNFGMMLQPEQLCWIISGAGTEKTSSEVLNAAKVLIIETQRCNSRYVDNL 420
DB 273 IQPVCLPRAGQALVDKICTVGTGNTQYVQAGVQLQEARVPIISNDVNGADFYGNQI 332
QY 421 TPAMI CAGFLQGNVDSQGDSSGGLV-----TSKNNIWLIGDTSWGSCKAKAYRPGYGN 476
DB 333 KPMFCAGYPEGGIDACQDSSGPPFCVDSISRTPRWLQGVISWGTGTCALQKPGVYTK 392
QY 477 VMVFTDWYKQMR 489
DB 393 VSDFFRWFQALK 405

RESULT 7

JC7731
membrane-bound arginine-specific serine proteinase precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 03-Feb-2003
C:Accession: JC7731; JC7775
R:Kishi, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.; Achauda,
J. Biochem. 130, 425-430, 2001
A:Title: Characterization of a membrane-bound arginine-specific serine protease from ra
A:Reference number: JC7731; MUID:21421307; PMID:11530019
A:Accession: JC7731
A:Molecule type: mRNA
A:Residues: 1-855 <KIS>
A:Cross-references: DBJ:AB049189
A:Experimental source: strain Male; 7-week-old
R:Satom, S.; Yamazaki, Y.; Tsuchiya, Y.; Tsuchiya, Y.; Hitomi, Y.; Iwanaga, T.; Fushiki, T.
Biochem. Biophys. Res. Commun. 287, 995-1002, 2001
A:Title: A role for membrane-type serine protease (MT-spl) in intestinal epithelial tur
A:Reference number: JC7775; PMID:11573963
A:Contents: Small intestine
A:Accession: JC7775
A:Molecule type: mRNA
A:Residues: 1-855 <SAT>
A:Cross-references: DBJ:AB037898
C:Comment: This enzyme, an epithelial-derived, type II integral membrane serine protease
of specific proteins or peptides on the brushborder membranes. It also participates in
lial migration and/or cell loss.
C:Genetics:
A:Gene: mt-spl

C:Superfamily: coagulation factor XI; trypsin homology
 C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:120-390/Product: plasma kallikrein heavy chain #status experimental <HCH>
 F:20-109/Domain: apple repeat <AP1>
 F:110-139/Domain: apple repeat <AP2>
 F:200-289/Domain: apple repeat <AP3>
 F:291-380/Domain: apple repeat <AP4>
 F:391-621/Domain: trypsin homology <TRY>
 F:21-104,47-77,51-55,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322
 F:127,215,308,396,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 19.8%; Score 538; DB 1; Length 638;
 Best Local Similarity 37.1%; Pred. No. 38-31;
 Matches 115; Conservative 52; Mismatches 117; Indels 26; Gaps 6;
 QY 185 CRMGYNPNFYSSQGVDSGTSFMKLTNSAGNVDIYKLYHSDACSSKAWSLRCLAC 244
 DB 340 CKEEGCKCSURLS---TDGSPRITVGMQSGYSLRLCKLVDPDCTTKI----- 387
 QY 245 GVNLNRSRRIYGGESALPGAWPMQVSLHVQNV---HVCQGSITPEWIVTAHCVKEKP 301
 DB 388 -----NARIYGGTNASLGFWQVSLQVQLVQTHLQCGSILGQWVLTAAHCFD-G 438
 QY 302 LNNPWHWTAPAGILROSFYGA-GYQVEKVIHSPNYDSKTKNDIALMKLQKPLTFNDL 360
 DB 439 IPYFDVRIYGGILSLSEITKETPSRSIKELIHQYKVSNGYDIALIKLQPLNTEF 498
 QY 361 VKPVLCPNPGMMLQPEQLCWISGWGATEKGTSEVLAANKVLLIETORCNSRVVDNLI 420
 DB 499 QKFIPLPSKADNTIYTNQVITGWTGKTEQGTQNLQKAIPLVFNESQKKI-RDYVI 557
 QY 421 TPAMICAGLQGVDSQCGSGLPLVTSKNNIWMILGTSWGSQCAKAYRPGVGNVMVF 480
 DB 558 NKQMICAGYKEGTDACKDGSGLPVCKHSGRWQLVIGTSWGSQGRKDPGVYTKVSEY 617
 QY 481 TDWIVQWRA 490
 DB 618 MDWILEKTQS 627

RESULT 10

KFHUI
 coagulation factor Xla (EC 3.4.21.27) precursor [validated] - human
 N:Alternate names: antihemophilic factor C; plasma thromboplastin antecedent
 C:Species: Homo sapiens (man)
 C:Date: 13-Aug-1986 #sequence revision 26-May-1994 #text_change 08-Dec-2000
 C:Accession: A27431; A00920; A37940
 R:Asakai, R.; Davie, E.W.; Chung, D.W.
 Biochemistry 26, 7221-7228, 1987
 A:Title: Organization of the gene for human factor XI.
 A:Reference number: A27431; MUID:88107663; PMID:2827746
 A:Accession: A27431
 A:Molecule type: DNA
 A:Residues: 1-625 <ASA>
 A:Cross-references: GB:M18295
 A>Note: the sequence shown follows the authors' translation
 R:Fujikawa, K.; Chung, D.W.; Hendrickson, L.E.; Davie, E.W.
 Biochemistry 25, 2417-2424, 1986
 A:Title: Amino acid sequence of human factor XI, a blood coagulation factor with four ty
 A:Reference number: A00920; MUID:86243360; PMID:3636155
 A:Accession: A00920
 A:Molecule type: mRNA
 A:Residues: 1-625 <FUJ>
 A:Cross-references: GB:M13142; NID:gl82832; PIDN:AAA52487.1; PID:gl82833
 R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.
 Biochemistry 30, 2056-2060, 1991
 A:Title: Location of the disulfide bonds in human coagulation factor XI: the presence of
 A:Reference number: A37940; MUID:91152017; PMID:1998667
 A:Accession: A37940
 A:Molecule type: protein

A:Residues: 28-33;35-49,'X',51-55,'X',57-63;70-75,'X',77-79;107-109,'X',111-112;132-139
 ;280-282,'X',284;285-297;313-316,'X',318-319;320-326,'X',328-330,'X',347-349;373,'X',37
 C:Comment: the proenzyme consists of two identical chains linked by one or more disulf
 he active site, and a heavy chain, which associates with high molecular weight (HMW) K1
 C:Genetics:
 A:Gene: GDB:F11
 A:Cross-references: GDB:119891; OMIM:264900
 A:Map position: 4q35-q35
 A:Introns: 19/1; 73/2; 109/1; 162/2; 199/1; 252/2; 289/1; 343/2; 379/1; 435/2; 494/1; 5
 C:Function:
 A:Description: catalyzes the proteolytic activation of coagulation factor IX
 A:Pathway: blood coagulation intrinsic pathway
 C:Superfamily: coagulation factor XI; trypsin homology
 C:Keywords: blood coagulation; duplication; glycoprotein; hemophilia C; homodimer; hydr
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-387/Product: coagulation factor Xla heavy chain #status experimental <HCH>
 F:109-108/Domain: apple repeat <AP1>
 F:109-198/Domain: apple repeat <AP2>
 F:199-288/Domain: apple repeat <AP3>
 F:290-379/Domain: apple repeat <AP4>
 F:388-625/Product: coagulation factor Xla light chain #status experimental <LCH>
 F:388-618/Domain: trypsin homology <TRY>
 F:20-103,514-581,571-599/Disulfide bonds: #status predicted
 F:29/Disulfide bonds: interchain #status experimental
 F:46-76,50-56,110-193,136-165,140-146,200-283,226-255,230-236,291-374,317-346,321-327,3
 F:90,126,353,450/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:339/Disulfide bonds: interchain #status predicted
 F:387-388/Cleavage site: Arg-11e (coagulation factor Xla) #status experimental
 F:431,480,575/Active site: His, Asp, Ser #status predicted
 F:491/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 19.6%; Score 533; DB 1; Length 625;

Best Local Similarity 30.5%; Pred. No. 6.8e-31;
 Matches 122; Conservative 63; Mismatches 129; Indels 86; Gaps 11;

QY 169 HPVC-----QDDWENYGRACRDMGYKNFYSSQGVDSGTSFMKLTNSAGNVDIY 222
 DB 233 HPGCLPFTFFSQEPKESQRLNLTSSGLPSTR--IKSKALSGLSQCRRHSIPVF 290
 QY 223 --KLYHS-----DACSKAVVSLRCI-----ACGVNL 248
 DB 291 CHSFFYHDTDFLGEELDIVAAKSHBACQLCTNAVCRCQFTYTPAQASQNEGKCYLKL 350
 QY 249 NSSR-----QSRIVGGESALPGAWPMQVSLHVQNV 278
 DB 351 SSNGSPKILHGRGIGSYTLRLCKMDNCTTKIKPRIVGGTASVRGEWPMQVLTHTSP 410
 QY 279 ---HVCQGSITPEWIVTAHCVKEKPLNNPWHWTAPAGILROSFM-----FYCAGYQVEK 330
 DB 411 TORHLCGSGIIGQWVLTAAHCF-YGVSEPKILRVYSGILNOSSEIKEDTSFFG----VQE 465
 QY 331 VISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPNFGMMLQPEQLCWISGWGATEBK 390
 DB 466 IIIHDQYKMAESGYDIALKLTETVNTYDSORPICLPSKEDENVIYTCWVTGWYRKL 525
 QY 391 GKTSEVLAANKVLLIETORCNSRVVDNLIITPAMICAGLQGVDSQCGSGLPLVTSKN 450
 DB 526 DKIQNTLQAKIPLVFNESQKEY-RGHKITHKVICAGYREGGKACKDGSGLPLSCRN 584
 QY 451 NIWMLIGTSWGSQCAKAYRPGVGNVMVFTDNIYRQMA 490
 DB 585 EVMLHVGITSGWGCQARERPGVYTNVVEYVDWILEKTQA 624

RESULT 11

S33777
 hepsin (EC 3.4.21.-) - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text_change 18-Jun-1999
 C:Accession: S33777; S32013
 R:Farley, D.; Reymond, F.; Nick, H.
 Biochim. Biophys. Acta 1173, 350-352, 1993
 A:Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase

A;Reference number: S33777; MUID:93305733; PMID:8318546

A;Accession: S33777

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-416 <PAR>

A;Cross-references: EMBL:X70900; NID:g57928; PIDN:CAA50256.1; PID:g57929

C;Superfamily: hepsin; trypsin homology

C;Keywords: hydrolase; liver; serine proteinase; transmembrane protein

F;22-44/Domain: transmembrane #status predicted <TMN>

F;162-399/Domain: trypsin homology <TRY>

F;187-203,290-358,321-337,348-380/Diulfide bonds: #status predicted

F;202,256,352/Active site: His, Asp, Ser #status predicted

Query Match

Best Local Similarity 19.5%; Score 529.5; DB 1; Length 416;

Matches 132; Conservative 66; Mismatches 177; Indels 57; Gaps 10;

QY 77 CTSKTKKALCITITLGTFLVGAALAAAGLLWKFMSKCSNSGTECDSSGTCINPSNWCQGV 136

Db 11 CCSRPKVA---ALTGTL-----FL-----TGGAASNAIVT 40

QY 137 SHCPGGEDENRCVRLGPNFLOVYSSQKSHVPCDDWNYGAAACRDMGYKNFY 196

Db 41 ILRSDDEPLYQVQLGPGDRLVLDKTEGTWRLCCSRNARVAGLCEENGFLRALAH 100

QY 197 SQGIVDDSGSTFWKLTSA-----GNVDIVYKLYHSDA---CSSKAVVSLRCIACGVN 247

Db 101 SELDVRTAGAN-----GTSGFVCDGGLPLAQRLLDVIVSVCDCPRGRFLATQCDCG-- 153

QY 248 LNSRSRRIYGGESALPGAPWQVSLHVQNVHVCVGSITPWIYTAHCEKPLNPNWH 307

Db 154 RRLPLVDRIYGGQDSSLGMPWQVSLRYDGLHCGSLSGDWLTAHCFPERNRVLSR 213

QY 308 WTAAGILRQSFYFAGYQVERKVI SHPNY-----DSKTKNDIALMKLQKPLTFNDLV 361

Db 214 WRVFAVAVATSP-HAVQLGVQAVIYHGGYLPFRDPTDENSNDIALVHLSPLLEYI 272

QY 362 KPVCPLNPGMLOPEGLCWISGWGATEEKGKTSVNLNAKULLIETQCNRYVDNLIT 421

Db 273 QPVCLPAAGALYDGVKVTCTGNGTQFYGOQAVLQEARVPIISNEVCNPFDFYGNQIK 332

QY 422 PAMICAGFLQGNVDSQCGSGGGLV----TSKNNTWLLIGTWSGGSCAKAYRPGVYGNV 477

Db 333 PKMFCAGYPEGIDACQSGSGHGFVCEDEISGTSRRLCGIVSGTGCALARKFGVYTKV 392

QY 478 MVFTDIYQMR 489

Db 393 IDPREWIFQAIA 404

RESULT 12

PLMS

Plasmin (EC 3.4.21.7) precursor - mouse

N;Contains: angiotensin; plasminogen

C;Species: Mus musculus (house mouse)

C;Date: 20-Sep-1991 #sequence revision 01-Nov-1996 #text_change 18-Jun-1999

C;Accession: A38514; S48202; S48203

R;Degen, S.J.P.; Bell, S.M.; Schaefer, L.A.; Elliott, R.W.

Genomics 8, 49-61, 1990

A;Title: Characterization of the cDNA coding for mouse plasminogen and localization of

A;Reference number: A38514; MUID:91184812; PMID:2081600

A;Accession: A38514

A;Molecule type: mRNA

A;Residues: 1-812 <DEG>

A;Cross-references: GB:J04766; NID:g200402; PIDN:AAA50168.1; PID:g200403

R;Liijnen, H.R.; van Hoef, B.; Beelen, V.; Collen, D.

Eur. J. Biochem. 224, 863-871, 1994

A;Title: Characterization of the murine plasma fibrinolytic system.

A;Reference number: S48202; MUID:95010076; PMID:7521120

A;Accession: S48202

A;Molecule type: protein

A;Residues: 20-25 <LIJ>

A;Accession: S48203

A;Molecule type: protein

A;Residues: 22-27 <LIJ>

C;Comment: Plasminogen is synthesized by the kidney and is present in plasma and many o

mediately after dissociation from the clot. In the presence of the inhibitor, the activ

e inhibitor, the activation involves also removal of the activation peptide.

C;Comment: Stromelysin 1 (see PIR:KCMSS1) acts on plasminogen to produce angiotatin. T

ecul in treating solid tumors.

C;Function:

A;Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a v

ns the walls of the graafian follicle; also activates the urokinase-type plasminogen ac

A;Pathway: fibrinolysis

C;Superfamily: plasmin; kringe homology; plasminogen-related protein precursor homolog

C;Keywords: angiotensin inhibitor; blood; duplication; fibrinolysis; glycoprotein; hyd

F;1-96/Domain: plasminogen-related protein precursor homology <PLPH>

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-812/Product: plasminogen #status predicted <PRO>

F;20-96/Domain: activation peptide #status predicted <APT>

F;79-466/Product: angiotensin #status predicted <AST>

F;97-581,582-812/Product: plasmin #status predicted <MAT>

F;97-581/Domain: chain A #status predicted <ACH>

F;103-181/Domain: kringe homology <KR1>

F;185-262/Domain: kringe homology <KR2>

F;275-352/Domain: kringe homology <KR3>

F;377-454/Domain: kringe homology <KR4>

F;481-560/Domain: kringe homology <KR5>

F;582-812/Domain: chain B #status predicted <BCH>

F;582-805/Domain: trypsin homology <TRY>

F;49-75,53-63,103-181,124-164,152-176,185-262,188-316,206-245,234-257,275-352,296-335,3

Bonds: #status predicted

F;78-79/Cleavage site: Glu-Asn (stromelysin 1) #status predicted

F;136,308/Binding site: carboxylate (Asn) (covalent) #status predicted

F;466-467/Cleavage site: Thr-Val (stromelysin 1) #status predicted

F;581-582/Cleavage site: Arg-Val (plasminogen activator) #status experimental

F;624,667,763/Active site: His, Asp, Ser #status predicted

Query Match 19.1%; Score 518.5; DB 1; Length 812;

Best Local Similarity 32.7%; Pred. No. 1e-29;

Matches 136; Conservative 54; Mismatches 143; Indels 83; Gaps 16;

QY 106 WKFMGSK-CSNSGIECDSSGTCI-----NPSNWCQGVSHC--PGGEDENRCVRLYXPN 155

Db 446 WEVCNLRKCSERG-----GSVVELPTVSPGSPSDSETDCMYGKDYRGKTAVTAAG 499

QY 156 FLQVYSSQKSHVPC-----QDDWNYGAAACRDMG-----YKKNFYSSQGIYVDDSG 205

Db 500 TPCQGWAAQEPHRSIFTPQTNPRADLEKNYCRNPDGVDVNGPWCYTTN----- 547

QY 206 STSPMKLNTSAGNVDIYKLYHSDACSKSAVSLRCIACGVNLNLSRQ--SRIVGGESAL 263

Db 548 -----PRKLY--DYCDIPLCASASSFECKGKQVPEKKCPGRVVGCVAN 599

QY 264 PGANPWQVSLHVQ--NVHVCVGSITPWIYTAHCEKPLNPNWHWTAFAGILRQSFMF 321

Db 590 PHSWQVSLRTRFTGQHFQCGTLIAPEWLVLTAAHCLERS-SRPEFYKVLGAHEE---- 644

QY 322 YGAGYQVR-----KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPNFGMLQPE 376

Db 645 YIRGLDVEISVAKLLEPN-----NRDIALKLSRPATITDKVIPACLPSPNYMVADR 698

QY 377 QLCWISGWGATE---EKGTSEVLNAKULLIETQCNRSRVVDNLITPAMICAGFLQGN 433

Db 699 TICITIGMETQGTFGAGRLKE---ACLPIENKVCNRYEYLNRRVKSTELCAGQLAGG 754

QY 434 VDSQCGSDGGPLVTSKKNINWLLIGTWSGGSCAKAYRPGVYGNVMTDVIYQMR 489

Db 755 VDSQCGSDGGPLVCFEKDKYILQGVTSWGLGCAKPNKPGVYVVRVSRFVDVIERMR 810

RESULT 13

S11674

acrosin (EC 3.4.21.10) precursor - human

C;Species: Homo sapiens (man)

C>Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 16-Jun-2000
C/Accession: S11674; S23499; S12063; A61022; S03330
R;Keime, S.; Adham, I.M.; Engel, W.
Eur. J. Biochem. 190, 195-200, 1990
A>Title: Nucleotide sequence and exon-intron organization of the human proacrosin gene.
A/Reference number: S11674; MUID:90306003; PMID:2114285
A/Accession: S11674
A/Molecule type: DNA
A/Residues: 1-421 <KEI>
A/Cross-references: EMBL:X54017; NID:935582; PIDN:CAA37964.1; PID:G1216165
A/Note: the authors translated the codon AGG for residue 64 as Thr and CTG for residue 2
R;Vazquez-Levin, M.H.; Reventosa, J.; Gordon, J.W.
Eur. J. Biochem. 207, 23-26, 1992
A>Title: Molecular cloning, sequencing and restriction mapping of the genomic sequence
A/Reference number: S23499; MUID:92331659; PMID:1628652
A/Accession: S23499
A/Molecule type: DNA
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-421 <VAZ>
A/Cross-references: EMBL:M77378
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992
R;Keime, S.
submitted to the EMBL Data Library, December 1989
A/Reference number: S12063
A/Accession: S12063
A/Molecule type: DNA
A/Residues: 1-225, 'R', 227-421 <KEI2>
A/Cross-references: EMBL:X54017
R;Adham, I.M.; Klemm, U.; Maier, W.M.; Engel, W.
Hum. Genet. 84, 125-128, 1990
A>Title: Molecular cloning of human proacrosin cDNA.
A/Reference number: A61022; MUID:90128988; PMID:2298447
A/Accession: A61022
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-63, 'T', 65-225, 'V', 227-267, 'R', 269-421 <ADH>
R;Baba, T.; Watanabe, K.; Kashiwabara, S.I.; Arai, Y.
FEBS Lett. 244, 296-300, 1989
A>Title: Primary structure of human proacrosin deduced from its cDNA sequence.
A/Reference number: S03330; MUID:89153568; PMID:2493394
A/Accession: S03330
A/Molecule type: mRNA
A/Residues: 1-63, 'T', 65-119, 'V', 121-165, 'L', 167-267, 'R', 269-344, 'R', 346-421 <BAB>
A/Cross-references: EMBL:Y00970; NID:928325; PIDN:CAA6784.1; PID:G28326
C/Genetics:
A/Gene: GDB:ACR
A/Cross-references: GDB:119645; OMIM:102480
A/Map position: 22q13-22qter
A/Introns: 26/2, 94/2, 189/2, 237/3
C/Superfamily: acrosin; trypsin homology
C/Keywords: glycoprotein; hydrolase; serine proteinase; sperm
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-421/Product: acrosin #status predicted <MAR>
F:20-42/Product: acrosin light chain #status predicted <LCH>
F:43-421/Product: acrosin heavy chain #status predicted <HCH>
F:43-285/Domain: trypsin homology <TRY>
F:302-379/Region: proline-rich
F:22-210/Binding site: carbonylate (Asn) (covalent) #status predicted
F:25-154/Disulfide bonds: #status predicted
F:29-162/Disulfide bonds: #status predicted
F:73-89/Disulfide bonds: #status predicted
F:88-142,240/Active site: His, Asp, Ser #status predicted
F:177-246/Disulfide bonds: #status predicted
F:209-225/Disulfide bonds: #status predicted
F:236-266/Disulfide bonds: #status predicted

Query Match 18.4%; Score 501; DB 1; Length 421;
Best Local Similarity 40.8%; Pred. No. 8.9e-25;
Matches 107; Conservative 42; Mismatches 87; Indels 26; Gaps 10;

QY 244 CQVNLNSRQS--RIVGGSALPGAWPQVSLHV-----QNVHVCQGSIIITPEWIVTAH 296
A34170
A34170
N:Alternate names: 53K fucose-binding protein
C/Species: Sus scrofa domestica (domestic pig)

QY 29 C-VEKPLNNPWHWTAFAGILRQSFMYG-----AGYQ---VEKVISHPNYDSKTKNNDI 346
Db 89 CFVCK--NNVHDLVFGAKE---ITYGNKPKAPLQBRVYVEKIIIEHKYNSATEGNDI 143
QY 347 ALMKLQKPLFNDLVKPVCLPN--PGMMLQPEQLCWISGNGATEEGK-KTSEVLNAKVL 404
Db 144 ALVEITPISCGRFIGCCLPHFKAGLPRGSGSCWVGWGYIEKAPRPSSTILMEARVDL 203
QY 405 IETQCRNSRVYVNDLITPAMICAGFLQGNVDSQSGSGGLV--TSKNINMWLIGDTSWG 462
Db 204 IDLLCNSTQWYNGRVQPTWVCAGYFVGKIDTCQDGGGLMKCKDSKESAYVVGITSWG 263
QY 463 SGCAKAYRPGVYGVNVVFTDWI 484
Db 264 VGCALAKRPGIYTATWPLNWI 285

RESULT 14
T30337
polyprotein - African clawed frog
C/Species: Xenopus laevis (African clawed frog)
C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 03-Feb-2003
C/Accession: T30337
R;fang, J.C.; Lindsay, L.L.; Hedrick, J.L.
submitted to the EMBL Data Library, March 1998
A/Description: cDNA cloning of ovochymase, a chymotrypsin-like protease released from X
A/Reference number: Z20829
A/Accession: T30337
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1524 <YAN>
A/Cross-references: EMBL:U81290; NID:G2981640; PID:G2981641; PIDN:AAC2471.1
C/Superfamily: trypsin related polyprotein; trypsin homology

Query Match 18.4%; Score 501; DB 2; Length 1524;
Best Local Similarity 33.9%; Pred. No. 4e-28;
Matches 124; Conservative 59; Mismatches 129; Indels 54; Gaps 13;

QY 134 DGVSHCGGSDENRCVRLYGNPFIQVSSQRKSWHPVCQDDWNNYGRACRDMGYKNN 193
Db 486 DAVEYDGAEEKLIARLCGYLPLPI--SSPENTWMLIRKTDWNSY-----PGPKVK 537
QY 194 F-----YSSGGIVDDSGSTFSMKLNTSAGNVDIYKLYHSDACSSKAVVSLRCIACG 245
Db 538 FSVFVPEKQFSLP--VDDTPTISM--LHPRAIALDV-----CG 571
QY 246 VNLNRR--QSRIVGGSALPGAWPQVSLHVQNVHVCQGSIIITPEWIVTAHCVKPLN 303
Db 572 MAPMTPKWLPRIVGGEASPSWPHQVQIFPLRTEHCEGALISQWILTAAHCIRAA-- 629
QY 304 NPWHWTAFAGILRQSFMYGAGYQVEKVIS---HPNYSKTKNNDIALMKLQKPLTFNDL 360
Db 630 EPSYVTVIAG--DHNRLNTESTEQIENIKTIHNDYNSYTDNDIALLYDEPLDLNDF 687
QY 361 VKPVCLPNQGMMLQPEQLCWISGNGATEEGKKTSEVLNAKVLIIETQCRNSRVYVNDLI 420
Db 688 VRPVCLPEPEVLTASCVVTVGNGTADGGPQALQLOQLPILDSIICNTSY--YSGEL 746
QY 421 TPAMI CAGFLQ--NVDSCQDGGGLV--TSKNINMWLIGDTSWGSGCAKAYRPGVYGVNV 478
Db 747 TDHMLCAGFPSSKEKDACQDGGGLVQCNEKEQFSIYGLVSWGEGCGRVSKEPGYITKVR 806
QY 479 VFTDWI 484
Db 807 LFTWI 812

RESULT 15
A34170
A34170
N:Alternate names: 53K fucose-binding protein
C/Species: Sus scrofa domestica (domestic pig)

[illegible]

; PRIOR FILING DATE: 2001-11-13
 ; PRIOR APPLICATION NUMBER: US 60/335,394
 ; PRIOR FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/332,464
 ; PRIOR FILING DATE: 2001-11-21
 ; PRIOR APPLICATION NUMBER: US 60/334,393
 ; PRIOR FILING DATE: 2001-11-29
 ; PRIOR APPLICATION NUMBER: US 60/340,376
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: US 60/347,211
 ; PRIOR FILING DATE: 2002-01-08
 ; PRIOR APPLICATION NUMBER: US 60/347,349
 ; PRIOR FILING DATE: 2002-01-10
 ; PRIOR APPLICATION NUMBER: US 60/355,250
 ; PRIOR FILING DATE: 2002-02-08
 ; PRIOR APPLICATION NUMBER: US 60/356,714
 ; PRIOR FILING DATE: 2002-02-13
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 1386
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 585
 ; LENGTH: 1479
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-295-027-585

Alignment Scores:
 Pred. No.: 3,186-307 Length: 1479
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 16 Gaps: 0

US-09-615-285B-2 (1-492) x US-10-295-027-585 (1-1479)

QY	1	MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr	20
DB	1	ATGGCTTTGAATCAGGTCACCAACAGCTTATGAGCTTACTATGAAACCACTGGATAC	60
QY	21	GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis	40
DB	61	CAACCGGAACCCCTATCCCGCACAGCCCACTGTGTCCCACTGTCTACGAGGTGCAT	120
QY	41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60
DB	121	CGGGCTCAGTACTACCCGTCGCCCGTGCACGACCCAGTACGCCCGAGGGTCTCTGACGACGCT	180
QY	61	SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys	80
DB	181	TCCAAACCCGTCGCTCTGCACGACGCCCAATCCCAATCCCGGACAGTGTGCACACTCANAG	240
QY	81	ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu	100
DB	241	ACTAAGAAAGCACTGTGCATCACTTGACCCCTGGGGACCTTCTCTGTTGGGAGCTGGGCTG	300
QY	101	AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys	120
DB	301	GCCTGTGCTTACTCTGGAAGTTATGGGACAGAGTCTCCAACTCTGGATAGAGTGC	360
QY	121	AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro	140
DB	361	GACTCCTCAGGTACCTGCATCAACCCCTTAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	420
QY	141	GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal	160
DB	421	GGCGGGAGGACGAGATCGGT	480
QY	161	TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr	180
DB	481	TACTCATCTCAGAGAACTCTGGCACCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	540
QY	181	GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle	200

DB	S41	GGGCGGGCGCTGCAGGACATGGGCTATAGAAATATTTTACTCTAGCCAGGAATA	600
QY	201	ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp	220
DB	601	GTGGATGACAGCGGATCCACGACTTTATGAAACTGAACAACAAGTCCGGCAATGTCTGAT	660
QY	221	IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg	240
DB	661	ATCTATATAAACTGTACACAGTGTCTCTCTTCAAAAGCAGTGTCTTTTACGC	720
QY	241	CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyLeu	260
DB	721	TGTATAGCTTGGGGGTCAACTTGAACCAAGCCGACAGCAGGATCTGGGCGCGAG	780
QY	261	SerAlaLeuProGlyAlaTrpProTyrGlnValSerLeuHisValGlnValHisVal	280
DB	781	AGCGGCTTCCCGGGGCTGGCCCTGGCAGTGTGCAGCTGCAGAACGTCACAGTG	840
QY	281	CysGlyGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGlyLys	300
DB	841	TGCGGAGCTCCATCATCCCCCGAGTGGATCGTACAGCCGCCACCTGCGTGGAAAAA	900
QY	301	ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet	320
DB	901	CCTCTTAACATCCATGCGCATTTGACGGCATTTTGGGGATTTTGAGACATCTTTCAAG	960
QY	321	PheTyrGlyAlaGlyTyrGlnValGlyLysValIleSerHisProAsnTyrAspSerLys	340
DB	961	TTCTATGAGCGGATACCAAGTAGAAAAAGTATTTCTCATCTCAATATGACTCCAAG	1020
QY	341	ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu	360
DB	1021	ACCAAGAACATGACATTTGCTGTGAGCTGCAGAGCTCTGACTTTTCAACGACCTA	1080
QY	361	ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp	380
DB	1081	GTGAACACAGTGTCTGCCCAACCCAGGATGATGCTGCAGCCAGACAGCTCTCTCTGG	1140
QY	381	IleSerGlyTrpGlyAlaThrGluGlyLysGlyLysThrSerGluValLeuAsnAlaAla	400
DB	1141	ATTTCCGGTGGGGGCGCCCGAGGAGAAAGGAGACCTCAGAGTGTGAACGCTGCC	1200
QY	401	LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle	420
DB	1201	AAGTGTCTTCTCATTTAGACACAGATGCAACAGCAGATATGCTTATGACACCTGATC	1260
QY	421	ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp	440
DB	1261	ACACAGCATGATCTGTGCGGCTTCTGACGGGAACTGCTGATTTTGGCAGGCTGAC	1320
QY	441	SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpIleGlyAspThrSer	460
DB	1321	AGTGGGGGCTCTGTGTCATCTCGAAGAACTATCTGGTGGCTGATGGGATACAAGC	1380
QY	461	TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe	480
DB	1381	TGGGTTCTGGCTGTGCCAAGCTTACAGACAGGAGTGTACGGGAATGTGTGTTATTC	1440
QY	481	ThrAspTrpLeuTyrArgGlnMetArgAlaAspGly	492
DB	1441	ACGAGCTGGATTTATCGACAAATGAGGCGACGCGC	1476

RESULT 2

US-09-981-353-22
 ; Sequence 22: Application US/09981353
 ; Patent No.: US20020160392A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lasek, Amy W.
 ; APPLICANT: Jones, David A.
 ; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
 ; FILE REFERENCE: PA-0038 US
 ; CURRENT APPLICATION NUMBER: US/09/981,353

; CURRENT FILING DATE: 2001-10-11
 ; NUMBER OF SEQ ID NOS: 194
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 22
 ; LENGTH: 2486
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Incyte ID No. US20020160392A1 991163CB1
 US-09-981-353-22

Alignment Scores:
 Pred. No.: 2,68e-306 Length: 2486
 Score: 2712.00 Matches: 491
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 99.82% Indels: 0
 DB: 9 Gaps: 0

US-09-615-285b-2 (1-492) x US-09-981-353-22 (1-2486)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 DB 149 ATGGCTTTGAATCAGGTCACCACTGCTATTGGACCTTACTATGAAAAACCATGGATAC 208
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValPProThrValTyrGluValHis 40
 DB 209 CAAACCGGAAACCCCTATCCCGACAGCCCACTGGTGGTCCCACTGCTACGAGGTGCAT 268
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProGValLeuThrGluAla 60
 DB 269 CCGGCTCAGTACTACCCGTCGCCCGTCCCACTACGCGCCGAGGGTCCCTGACACAGGCT 328
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 DB 329 TCCAAACCCGTCGTCGACGAGCCCAATCCCACTCCGAGACAGTGTGCACCTCAAG 388
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 DB 389 ACTAAGAAAGACATGGCATCACTTACCTGGGACCTTCCTCGTGGGAGCTGGCGTG 448
 QY 101 AlaAlaGlyLeuLeuThrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 DB 449 GCGGCTGGCTACTCTGGAAGTTTCATGGCAGCAAGTCTCCAACTCTGGGATAGAGTGC 508
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 DB 509 GACTCTCAGGTACTTCGCATCAACCCCTCACTGCTGTAAGTGGTGTGCACATGCCCC 568
 QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPhelIleLeuGlnVal 160
 DB 569 GCGGGGAGGAGAGAAATCGGTGTGTTCGCTCTACGGACCAAACTTCATCCTTCAGGTG 628
 QY 161 TyrSerSerGlnArgLysSerThrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
 DB 629 TACTCATCTCAGAGAGAGTCTGGCACCTGTGGCCAAAGCAGCTGGAACGAGAACTAC 688
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPhetYrSerSerGlnGlyIle 200
 DB 689 GCGCGCGCGCTCGCAGGACATGGCTATAGAAATATTTTACTCTAGCAAGGAATA 748
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 DB 749 GTGGATGACAGCGGATCCCAAGCTTTTATGAACATGAACACAACTGGCGGCATGTGAT 808
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 DB 809 ATCTATAAAAACTGTACACAGTATGCTGTCTTCAAAAGCAGTGGTGTCTTTACGC 868
 QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 DB 869 TGTATAGCTTGGGGGTCACTTGAATCTCAAGCCGCCAGCAGCAATGTGGCGGCGAG 928

QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 DB 929 AGCGGCTTCCCGGGGCTGGCCCTGGCAGTACGCTGCACGCTCCAGAACGTCACGCTG 988
 QY 281 CysGlyGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
 DB 989 TCGGAGGCTCCATCATCACCCCGAGTGCACGCTGACAGCCGCCACTCGGTGGAAAAA 1048
 QY 301 ProLeuAsnAsnProTyrHisThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 DB 1049 CCTCTTAACATCCATGGCATTGGACCGCATTTGGGGGATTTTGAGACATCTTTCTATG 1108
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 DB 1109 TTCTATGGAGCGGATACCAAGTAGAAAAAGTATTCTCATCAAAATTATGATCTCAAG 1168
 QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 DB 1169 ACCAAGAACATGACATTTGGCTGATGAAGCTGCAGAGCTCTGACTTTCAACGACCTA 1228
 QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
 DB 1229 GTGAAACCAAGTGTCTGTGCCCAACCCAGGATGATGCTGCAGCCAGACAGCTCTGCTG 1288
 QY 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 DB 1289 ATTTCCGGGTGGGGGGCCACCGAGGAGAAAGGAGACCTCAGAGTGTGACGCTGCC 1348
 QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 DB 1349 AAGTGTCTTCTCATTTGAGACACAGATGCAACAGCAGATATGTCTATGACACCTGATC 1408
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 DB 1409 ACACGACGATGATCTGTGCCGCTTCTCGCAGGGAAACGTCGATTTCTGCCAGGCTGAC 1468
 QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrPheLeuIleGlyAspThrSer 460
 DB 1469 AGTGGAGGGCTCTGGTCACTTCGAAAGAACATATCTGGTGGCTGATAGGGGATACAGC 1528
 QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 DB 1529 TGGGGTCTGGCTGTGCCAAAGCTTACAGACAGGAGTGTACGGGAATGTGATGTATTTC 1588
 QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
 DB 1589 ACGGACTGGATTATCGAACAAATGAGGCGAGACGGC 1624

RESULT 3

US-10-116-802-203
 ; Sequence 203, Application US/10116802
 ; Publication No. US20030065157A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Amy Lasek
 ; TITLE OF INVENTION: GENES EXPRESSED IN LUNG CANCER
 ; FILE REFERENCE: PA-0045 US
 ; CURRENT APPLICATION NUMBER: US/10/116,802
 ; CURRENT FILING DATE: 2002-04-04
 ; PRIOR APPLICATION NUMBER: 60/281,593
 ; PRIOR FILING DATE: 2001-04-04
 ; NUMBER OF SEQ ID NOS: 519
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 203
 ; LENGTH: 2486
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No: 991163CB1
 US-10-116-802-203

Alignment Scores:

Pred. No.: 2,68e-306 Length: 2486
 Score: 2712.00 Matches: 491
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 99.82% Indels: 0
 DB: 13 Gaps: 0

US-09-615-285B-2 (1-492) x US-10-116-802-203 (1-2486)

```

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 149 ATGGCTTTGAACACAGGGTACACACCTGCTATTGACCTTACTATGAACACCTATGGATAC 208
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
DB 209 CAACCGGAAACCCCTATCCGACACAGCCACTGTGGTCCCACTGCTACGAGGTGCAI 268
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 269 CCGGCTCAGTACTACCCGCTCCCGCTGGCCAGTACGCCCGGAGGCTCCTGACGAGGCT 328
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 329 TCCAAACCCCGTCTGTGACGAGCCCAATCCCACTCCGGACAGGTGTGCACCTCAAG 388
QY 81 ThrLysLeuAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB 389 ACTAAGAAAGCACTGTGCATCACCTTGACCTTGGGACCTTCTCTGGTGGAGCTGCGCTG 448
QY 101 AlaAlaClyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB 449 GCGCTGGGCTACTCTGGAAGTTTCATGGGACAGAGTCTCCAACTCTGGGATAGAGTGC 508
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
DB 509 GACTCTCTAGGTACCTGCATCAACCCCTTAACTGGTGTGATGGCGTGTACACTGCCCC 568
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
DB 569 GCGGGGAGGAGCAGAGATCGGTGTCTTCCCTCTACGGACCAAACTTCACTCTCAGGTG 628
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
DB 629 TACTCATCTCAGAGAGAGTCTTGGACCCCTGTGGCCAGACGACTGGACGAGACTAC 688
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
DB 689 GGGCGGGCGGCTCGAGGAGACATGGGCTATAAGAAATATTTTACTCTAGCCCAAGGAATA 748
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB 749 GTGGATGACAGCGGATCCACAGCTTATGAACCTGAACCAACTGCGGCAATGTCAT 808
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 809 ATCTATAAAAAAAGCTGTACCACAGTATGCTGTCTTCAAAAGCAGTGGTTCCTTTACGC 868
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
DB 869 TGTATAGCTGTGGGGTCAACTTGAACCTCAGCGCCAGACAGAGATTGTGGCGGCGAG 928
QY 261 SerAlaLeuProGlyAlaTyrProThrGlnValSerLeuHisValGlnAsnValHisVal 280
DB 929 AGCGCGCTCCCGGGGCTGGGCTGTGGAGGTGACGCTGACGCTCCAGAAAGCTCCACGCTG 988
QY 281 CysGlyGlySerIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
DB 989 TGGGAGGCTTCATCATCAACCCCGAGTGGACCGGTGACAGCGGCCCACTGGTGGAAAAA 1048
QY 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
DB 1049 CCTCTTAAACATCCATGGCATTTGGAGCGCATTTGGGGGATTTTGGAGCAATCTTTTCATG 1108

```

```

QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
DB 1109 TTCTATGAGCGCGGATACCAAGTAGAAAAAGTGAATTTCTCATCCCAATATGACTCCAAAG 1168
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
DB 1169 ACCAAGAACAATGACATTGCGCTGATGAAGCTGCAGAACCTCTGACTTTTCAACGACCTA 1228
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
DB 1229 GTGAACACAGTGTGTGCCCCAACCCAGGATGATCTGCAGCGAAGACAGCTCTGCTGG 1288
QY 381 IleSerGlyTrpGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
DB 1289 ATTTCCGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAACTGTGAACGCTGCC 1348
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
DB 1349 AAGTGTCTTCTATTGAGACACAGATGCAACAGCAGATATGCTATGACAACTGATC 1408
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
DB 1409 ACACGACCATGATCTGTGCGGCTTCTTCAGGGGAGAGCTGATTTCTTCCAGCGGTGAC 1468
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
DB 1469 AGTGGAGGGCTCTGGTCACTTCGAAGAACATATCTGTGCTGATAGCGGATACAGC 1528
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
DB 1529 TGGGGTCTGGCTGTGCCAAAGCTTACAGCAGGAGTGTACGGGAATGTGATGTTATTC 1588
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
DB 1589 ACGGACTGGATTTATCGACAAATGAGGGGAGACGGC 1624

```

RESULT 4

```

US-10-205-823-414
; Sequence 414, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endegs, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Monsey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 414
; LENGTH: 3226
; TYPE: DNA

```


US-09-759-143-931

Alignment Scores:

Pred. No.: 2,736-306 Length: 1476
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 9 Gaps: 0

US-09-615-285B-2 (1-492) x US-09-759-143-931 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 1 ATGGCTTTTGAACCTCAGGGTCCACACAGCTATTGACCTTACTATGAAACCATGGATAC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValProThrValHis 40
DB 61 CAACCGGAAACCCCTATCCGACAGCCCACTGGTCCCACTGCTACGAGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CCGGCTCAGTACATACCCGCTCCCGCTGCCAGTACGCCCGGAGGGTCTTGACCGGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 181 TCAACCCCGCTGCTGTCGACGAGCCCAATCCCAATCCCGGACAGTGTGCACCTCAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB 241 ACTAAGAAAGCACTGTGCATCACCTTGACCTCCCTGGGGACCTTCCCTGGTGGAGCTGCCTG 300
QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB 301 GCCGCTGGCTACTCTGGAAGTTTATGGCAGCAAGTCTCCAACTCTGGGATAGATGTC 360
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
DB 361 GACTCCTCAGGTACCTGTCATCAACCCCTCAACTGGTGTGATGGCTGCACACTGCCCC 420
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
DB 421 GCGGGGAGGAGGAGATCGGTGTGCTTCGCTCTACGATCAAACTTCACTTCCTCAGGTG 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
DB 481 TACTCATCTCAGAGAAAGTCTGGCACCCTGTGTGCCAAGACGACTGGAACGAGAACTAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
DB 541 GCGCGCGGCTGTCAGGAGCATGGCTATAAGATAATTTTACTCTAGCCAGGAATA 600
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB 601 GTGGATGACAGCGGATCCACCAAGCTTATGAACCTGAACACAAAGTCCCGCAATGTCGAT 660
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 ATCTATAAAACTGTACACAGTATGCTGTCTTCARAAAGCAGTGGTTCTTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
DB 721 TGTATAGCTTGGGGTCAACTTGAACCTCAAGCCCGCAGACGAGATTTGGGGCGCGAG 780
QY 261 SerAlaLeuProGlyValAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
DB 781 AGCGGCTCCCGGGGCTGGCCCTGGCAGGTGTCAGCTGCAGTCCAGAACGTCACGTCG 840
QY 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
DB 841 TCGGAGGCTCCATCATCCCCCGAGTGTGTCGACAGCCCGCCCTCGTGGGAAAAA 900
QY 301 ProLeuAsnAsnProTyrTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320

DB 901 CCTTTAACAATCCATGCGCATTTGCGGGGATTTTGACACATCTTTTCATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValLysSerHisProAsnTyrAspSerLys 340
DB 961 TTCTATGAGCGGATACCAAGTAGAAAAGTATTTCTCATCCAAATATGACTCCAAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
DB 1021 ACCAAGAACAATGACATTTGCGCTGATGAAGCTGCAGAAAGCTCTGACTTTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
DB 1081 GTGAACACAGTGTCTGTGCCCAACCCAGGATGATGTCGAGCCAGAACACTCTGCTGG 1140
QY 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
DB 1141 ATTTCCGGGTGGGGGCCACCGAGGAGAAAGGAAACCTCAGAAGTGTCTGAACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
DB 1201 AAGGTGCTTCTCATTTGACACACAGATGCACAGCAGATATGCTATGACAACTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
DB 1261 ACACAGCAGCATGATCTGTGCGGCTTCTCTGCAGGGGAACGTCGATTTCTTCCAGGGTGC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
DB 1321 AGTGAGGGCTCTGTGTCACCTCCGAAGAACAATATCTGGTGGCTGATAGGGGATACAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
DB 1381 TGGGGTCTGCTGTGTCGCAAGCTTACAGACAGGAGTGTACGCGAATGTGATGGTATTC 1440
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
DB 1441 ACGGACTGGAATTTATCGACAAATGAGGCGACAGCGC 1476

RESULT 6

US-09-780-669-931
; Sequence 931, Application US/09780669
; Patent No. US2002005197A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Carter, Darriack
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C24
; CURRENT APPLICATION NUMBER: US/09/780.669
; CURRENT FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 943
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 931
; LENGTH: 1476

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-780-669-931

Alignment Scores:
  2,73e-306      Length: 1476
  Pred. No.:    2709.00    Matches: 491
  Score:        99.80%     Conservative: 0
  Percent Similarity: 99.80% Mismatches: 1
  Best Local Similarity: 99.71% Indels: 0
  Query Match: 99.71%     Gaps: 0
  DB: 9

US-09-615-285B-2 (1-492) x US-09-780-669-931 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 1 ATGGCTTTGAACTCAGGCTCACCACGCTATTGGACCTTACTATGAAACCATGGATAC 60

QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValProThrValHis 40
DB 61 CAACCGGAACCCCTATCCCGCAGCCACTGTGGTCCCACTGTCTACGAGGTGCAT 120

QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CGGGCTCAGTACTACCGCTCCCGCTGCCAGTACGCCCGGAGGGTCTTACGACGAGCT 180

QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 181 TCCAAACCCGCTGCTGTCAGGAGGCCCAATCCCCATCCGGGACAGGTGACCTCAAG 240

QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB 241 ACTAAGAAGACACTGTGCATCACCTTGACCTTGGGGACCTTCTCGTGGGAGCTGCGCTG 300

QY 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB 301 GCGCTGGCGTACTCTGGAAGTTCTAGGACAGCTGCTCAACTCTGGGATAGAGTGC 360

QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
DB 361 GACTCTCAGGTACCTGCATCAACCCCTTAACTGGTGTGATGGCGTGCACACTGCCCC 420

QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleGlnVal 160
DB 421 GCGCGGGAGGACGAGAATCGGTGTCTCGCTCTACGGATCAAACTTCTCCTTCAGGTG 480

QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
DB 481 TACTCATCTCAGAGGAAGTCTTGGCACCCCTGTGTCAGAGCACTGGACGAGAACTAC 540

QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyTle 200
DB 541 GGGCGGGCGCTCGCAGGACATGGCTATAGAATAATTTTACTCTTAGCCAAAGGAATA 600

QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB 601 GTGGATGACAGGGGATCCACCAAGCTTTATGAAACTTGAAACACCAAGTCCGGGCAATGCAAT 660

QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 ATCTATAAAAACTGTACACAGTGTGCTGTCTTCTCAAAAGCAGTGTGTTCTTTACGC 720

QY 241 CysIleAlaCysGlyValAsnLeuAsnSerArgGlnSerArgIleValGlyGlyGlu 260
DB 721 TGTATAGCTGGGGGTCAACTGAACTCAAGCCGCGCAGAGGATGTGGCGCGCAG 780

QY 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
DB 781 AGCGCGCTCCCGGGGCTTGGCCCTGGCAGGTGAGCTGCACGTCCAGAACTCCACGTG 840

QY 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
DB 841 TGGGAGGCTCCATCATCAACCCCGAGTGGATCGTGACAGCCGCCCTGCGGTGAAAAA 900

```

RESULT 7

```

US-09-822-827-931
; Sequence 931, Application US/09822827
; Patent No. US20020081680A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C1
; CURRENT APPLICATION NUMBER: US/09/822,827
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 931
; LENGTH: 1476
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-827-931

Alignment Scores:
  2,73e-306      Length: 1476
  Pred. No.:    2709.00    Matches: 491
  Score:        99.80%     Conservative: 0
  Percent Similarity: 99.80% Mismatches: 1
  Best Local Similarity: 99.71% Indels: 0
  Query Match: 99.71%     Gaps: 0
  DB: 9

US-09-615-285B-2 (1-492) x US-09-822-827-931 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 1 ATGGCTTTGAACTCAGGCTCACCACGCTATTGGACCTTACTATGAAACCATGGATAC 60

```

QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValProThrValTyrGluValHis 40
 Db 61 CAACCGGAAACCCCTATCCCGACAGAGCCACTGGGTCCTCTACGAGGTGCAT 120
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 121 CCGGCTCAGTACTACCGCTCCCGTGGCCCGAGTACGCCCGAGGGTCTTGACGAGCT 180
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 181 TCAACCCCGCTCGTGTGACGAGCCCAATCCCGCATCCGGGACAGTGTGCACCTCAAG 240
 QY 81 ThrLysLysAlaLeuCysLeuThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 Db 241 ACTAAGAAAGCACTGTGCATCACTTACCTTGGGACCTTCTCGTGGGAGCTGGCTG 300
 QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 301 GCGCTGGGCTACTCTGGAGTTCATGGGCGAGAGTCTCCAACTTGGGATAGAGTGC 360
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 Db 361 GACTCCTCAGGTACCTGCATCAACCCCTCAACTTGGTGTGATGGCGGTGCACACTGCC 420
 QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 421 GCGGGGAGGAGGAGGAGTGGTGTTCGCTCTACGATCAAACTTCATCTCTCAGGTG 480
 QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 Db 481 TACTCATCTCAGAGAGTCTGGCACTCTGGTGGCAAGACGACTGGAAAGAGAACTAC 540
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 Db 541 GGGCGGGGGCGCTGCAGGAGACATGGGCTATAGAGTAATTTTACTCTAGCCAGGAA 600
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 601 GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACAGAGTGGCGGCAATGTGAT 660
 QY 221 IleTyrLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Db 661 ATCTATAAAACTGTACACAGTATGCTGCTCTTCAAAAGCAGTGGTTCCTTACGC 720
 QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 Db 721 TGTATAGCTGCGGGTCAACTTTGAACCTCAAGCGCCAGACAGAGTGTGGCGGCGAG 780
 QY 261 SerAlaLeuProGlyAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 Db 781 AGCGGCTCCCGGGGCTGGCCCTGGCAGTTCAGCTGCAGTCCAGAAAGTCCAGCTG 840
 QY 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
 Db 841 TCGGAGGCTCCATCATCCCCCGAGTGGATCGTGACAGCGGCCACTGCTGGTGA 900
 QY 301 ProLeuAsnAspProTyrPheTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 Db 901 CTCTTAACTATCCATGGCATGGACGCAATTTGGGGATTTTGAGACAACTTTTCATG 960
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 Db 961 TTTCTATGGAGCGGATCCAGTAGAAAAAGTATTTCTCATCCAAATTTAGTCTCAAG 1020
 QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 Db 1021 ACCAAGAAACAATGACATTCGCTGTATGAGTGCAGAGCCCTCTGATTTCAACAGACCTA 1080
 QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
 Db 1081 GTGAACCAAGTGTGTCTGCCCAACCCAGGATGATGCTGCAGCCAGAACAGCTCTGTGG 1140

QY 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyIleThrSerGluValLeuAsnAlaAla 400
 Db 1141 ARTTCGGGTGGGGGCCACCGAGAGAAAGGAGACCTCAGAAGTGTGACACGTGCC 1200
 QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAsnLeuLeu 420
 Db 1201 AAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGCTATGACAACTGATC 1260
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGluGlyAsnValAspSerCysGlnGlyAsp 440
 Db 1261 ACACAGGCACTGATCTGCGCGTCTCTCAGGGGAACCTGATTTCTCCAGGGTGAC 1320
 QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
 Db 1321 AGTGAGGCGCTCTGGTCACTTCGAGAAACAATATCTGTGGTGTATAGGGGATACAAAGC 1380
 QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 Db 1381 TGGGTTCTGGCTGTGCCAAGCTTACAGACAGGAGTGTACGGGAATGTGATGGTATTC 1440
 QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
 Db 1441 ACGGACTGGATTTATCGACAAATGAGGGCGACGCGC 1476

RESULT 8

US-09-895-793-931
 ; Sequence 931, Application US/09895793
 ; Publication No. US20020192763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Jiang, Yuciu
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Ketter, Marc W.
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Day, Craig H.
 ; APPLICANT: Vedwick, Thomas S.
 ; APPLICANT: Carter, Darick
 ; APPLICANT: Li, Samuel X.
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Repler, William T.
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Hural, John
 ; APPLICANT: McNeill, Patricia D.
 ; APPLICANT: Houghton, Raymond L.
 ; APPLICANT: Vinals de Bassols, Carlota
 ; APPLICANT: FOV, Teresa
 ; APPLICANT: Fanger, Gary R.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE REFERENCE: 210121.534C2
 ; CURRENT APPLICATION NUMBER: US/09/895,793
 ; CURRENT FILING DATE: 2001-06-29
 ; NUMBER OF SEQ ID NOS: 982
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 931
 ; LENGTH: 1476
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-895-793-931

Alignment Scores:
 Pred. No.: 2,73e-306 Length: 1476
 Score: 2709.00 Matches: 491
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 99.71% Indels: 0
 DB: 9 Gaps: 0

US-09-615-285B-2 (1-492) x US-09-895-793-931 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 1 ATGGCTTTTGAACATCAGGTCACACAGCTATTGGACCTTACTATGAAACCATGZATAC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
Db 61 CAAACCGAAACCCCTATCCCGCACAGCCCACTGTGTCCCACTGTCTACAGAGTGCAT 120
QY 41 ProAlaGlyTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACTACCCGTCGCCCGTGGCCAGTACGCGCCGAGGCTCCTGACGAGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCGCTGCTGTCAGCAGACCCCAATCCCAATCCGGGACAGTGTGACCTCAAG 240
QY 81 ThrLysGlyAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 241 ACTAAGAAAGCACTGTGTCATCACTTGACCTGGGAGCTTCTCCTCGTGGAGCTGCGCTG 300
QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCGGCTGGCTACTCTGGAAGTTCTATGGGACAGAGTCTCCAACTCTGGGATAGAGTGC 360
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
Db 361 GACTCTCAGTACCTGTCATCAACCCCTTAACCTGTGTGATGGCGTGTCACTGCCCC 420
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 421 GCGGGGAGCAGAGAACTGGTGTGTTCGCTCTACGGATCAAACTTCATCCTTCAGGTG 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Db 481 TACTCATCTCAGAGGAAGTCTTGGACCCCTGTGTGCCAAGACGACTGCGAAGCAGAACTAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 541 GCGGGGCGGCTGTGAGGACATGGCTATAGGATATATTTTACTTACCCAGGAGATA 600
QY 201 ValAspAspSerGlySerThrPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 601 GTGGATGACAGCGGATCCACAGCTTTATGAACTGGAACACAGTGGCGGCAATGTCTGAT 660
QY 221 IleTyrLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATAAAACTGTATACAGATGATGCTGTCTTCAAAGCAGTGGTTCCTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGly 260
Db 721 TGTATAGCTGGGGGTCAACTTGAATCAAGCGCCAGAGCAGGATTTGTGGCGCGAG 780
QY 261 SerAlaLeuProGlyAlaTyrProThrGlnValSerLeuHisValGlnAsnValHisVal 280
Db 781 ACGCGCTCCCGGGGCTGGCCCTGGCAGTGCAGCTGCAGCTCCAGACGTCACAGTG 840
QY 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaHisCysValGluLys 300
Db 841 TCGGAGGCTCCATCATCCCGCGAGTGGATCGTGACAGCGCGCCCACTGCGTGGAAAA 900
QY 301 ProLeuAsnAsnProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTTAACATCCATGGCATTTGGACGGCATTTGGGGGATTTTGAGCAATCTTTCATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 961 TTCTATGGAGCGGATAGCAAGTAAAGAAAGTGAATTTCTCATCCAAATTTATGACTCCAAAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACATGACATTTGGCTGATGAAGCTGAGAGGCTCTGACTTTTCAACGACCTA 1080

QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1081 GTGAACACCATGTGTCTGCCCAACCCAGGATGATGTGTGCAGCCAGAACAGCTCTGTCTGG 1140
QY 381 IleSerGlyTyrGlyValThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCCGGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAGTGTGAAAGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1201 AAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGTCTATGACACACCTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACGACGATCATGTTGTGCGGCTTCTGTGCGGGACGTCGATTTCTTGCACGGGTGAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1321 AGTGGAGGCGCTCTGTCTCACTTCGAAGAAACAATATCTGTGTGCTGATAGGGGATACAGC 1380
QY 461 TrpGlySerGlyCysAlaAlaValAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGGTTCTGGCTGTGCCAAAGCTTACAGACAGAGAGTGTACGGGAATGTGATGTTATTC 1440
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1441 ACGGACTGGATTTATCGCAAAATGAGGCGACAGCGC 1476

RESULT 9

US-09-895-814-931
; Sequence 931, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yudi
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassola, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 931
; LENGTH: 1476
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-895-814-931

Alignment Scores:
Pred. No.: 2,73e-306 Length: 1476
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1

Query Match:	99.71%	Indels:	0
DB:	9	Gaps:	0
US-09-615-285B-2 (1-492) x US-09-895-814-931 (1-1476)			
QY	1	MetAlaLeuAsnSerGlySerProProAlaIleIleGlyProThrTyR-TyR-GluAsnHisGlyTyR	20
DB	1	ATGGCTTTGAATCAGAGGTCAACACAGCTATTGGACCTTACTATGAAACCAATGGATAC	60
QY	21	GlnProGluAsnProTyRProAlaGlnProThrValValProThrValTyRGluValHis	40
DB	61	CAACCGGAAACCCCTATCCCGCACGCCACACTGTGGTCCCACTGTCTACGAGTGCGAT	120
QY	41	ProAlaGlnTyR-TyRProSerProValProGlnTyR-AlaProArgValLeuThrGlnAla	60
DB	121	CGGGCTCAGTACTACCCGTCGCCGTCGCCCGAGGCTCTGAGCGAGGCT	180
QY	61	SerAsnProValValCysThrGlnProIlySerProSerGlyThrValCysThrSerIlys	80
DB	181	TCCAAACCCCGTCGTGTCACGACGCCAAATCCCCCATCCGGGACAGTGTGCACTCCAAG	240
QY	81	ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThr-PhelLeuValGlyAlaAlaLeu	100
DB	241	ACTAAGAAGCACTGTGCATCACCTTGACCTGGGAGACCTTCCTCGTGGAGCTGGCGTG	300
QY	101	AlaAlaGlyLeuLeuTrpLysPheMetGlySerIysCysSerAsnSerGlyIleGluCys	120
DB	301	GGCGTGGCCCTACTCTGGAATTCATGGGACGACAGTGTCTCAACTCTGGGATAGAGTGC	360
QY	121	AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro	140
DB	361	GACTCTCTCAGGTACCTGTCATCAACCCCTCTAACTGGTGTGATGGGTGTGCACACTGCCCC	420
QY	141	GlyGlyGluAspGluAsnArgCysValArgLeuTyR-GlyProAsnPheIleIleuGlnVal	160
DB	421	GGCGGGGAGGACGAAATCGGTGTGTGGCTCTACGGATCAAACTTCACTCCTCAGGTG	480
QY	161	TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyR	180
DB	481	TACTCATCTCAGAGGAAGTCTCTGGCACCCCTGTGTGCCAAGACGACTGGAACGAGAATCAC	540
QY	181	GlyArgAlaAlaCysArgAspMetGlyTyRLeuAsnAsnPheTyR-SerSerGlnGlyIle	200
DB	541	GGCGGGGGCGGCTCGCAGGGAATCGGGCTAAGAATTAATTTTACTCTAGCCAGAGGAATA	600
QY	201	ValAspAspSerGlySerThrSerPheMetLysLeuAsnThr-SerAlaGlyAsnValAsp	220
DB	601	GTGGATGACAGCGGATCCACAGCTTTTATGAAACTGAACACAAGTGGCGGCAATGTCGAT	660
QY	221	IleTyRLeuLysLeuTyRHisSerAspAlaCysSer-SerIysAlaValValSerLeuArg	240
DB	661	ACTCTATAAAAACCTGTACACAGTGATGCTGTTCCTCAAAGCAGTGGTTCCTTACGC	720
QY	241	CysIleAlaCysGlyValAsnLeuAsnSer-SerArgGlnSerArgIleValGlyGlyGlu	260
DB	721	TGTATAGCTGGCGGGTCAACTTCAACTCAAGCGCCACAGCAGGAGGATTGTGGCGCGCAG	780
QY	261	SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal	280
DB	781	AGCGGCTCCCGGGGGCTTGGCCCTGGCAGGTGAGCTGCGAGTCCAGAACTCCACGCTG	840
QY	281	CysGlyGlySerIleIleThrProGluTrpIleValThr-AlaAlaHisCysValGluLys	300
DB	841	TGGGAGGCTCCATCATCACCCCGAGTGGATCGTACAGCGCGCCCACTGGGTGGAAAAA	900
QY	301	ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet	320
DB	901	CTCTTTACATATCCATGTCATGGACGGCATTTGGGGGATTTTGGACATATCTTTCATG	960
QY	321	PheTyRAlaGlyTyRGlnValGlyValIleSerHisProAsnTyRAspSerIlys	340
DB	961	TTCTATAGGCGCGATACCAAGTAGAAAAAGTGATTTCTCATCCAAATATTGATCTCCAAG	1020

Alignment Scores:

Pred. No.: 2,73e-306 Length: 1476
 Score: 2709.00 Matches: 491
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 99.71% Indels: 0
 DB: 14 Gaps: 0

US-09-615-285B-2 (1-492) x US-10-012-896-931 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 DB 1 ATGGCTTTGACCTCAGGTCACCAACAGCTATGTGACCTTACTATGAAMACCATGGATAC 60
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 DB 61 CAACCGGAAACCCCTATCCGACAGCCCACTGTGGTCCCACTGTCTACGAGGTGCAT 120
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 DB 121 CCGGCTCAGTACTACCCGCTCCCGCTGCCAGTACGCCGAGGGTCTCTGACGAGCT 180
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 DB 181 TCCARCCCGCTGCTGCGACGAGCCCAATCCCATCCGACAGTGTGACCATCTCAAG 240
 QY 81 ThrLysLeuAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyValAlaLeu 100
 DB 241 ACTAAGAAACACTGTGTCATCACTTGACCTGGGACCTTCTCTCGTGGAGCTGCCCTG 300
 QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 DB 301 GCGCTGGGCTACTCTGGAAGTTCATGGGACAGTGTCTCAACTCTGGGATAGATGC 360
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 DB 361 GACTCTCAGTACCTGTCATCAACCCCTCAACTGCTGTGATGGCTGCACACTGCCCC 420
 QY 141 GlyGlyGluAsnGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 DB 421 GCGGGGAGACAGAAATCGGTGTGTGCTCTACAGGATCAAACTTCATCTTCAGGTG 480
 QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 DB 481 TACTCATCTCAGAGGAAGTCTGCGACCTGTGTGCCAAGACGACTGGAAACGAGAACTAC 540
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 DB 541 GCGCGGCGCTGTCAGGACATGGCTATAGAAATAATTTTACTCTAGCCAGGAAATA 600
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 DB 601 GTGGATGACAGGGGATCCACCAAGCTTTATGAACCTGAAACACAACTGCGCGCAATGTCAT 660
 QY 221 IleTyrLysLeuTyrHisSerSerPalaCysSerSerLysAlaValValSerLeuArg 240
 DB 661 ATCTATAAAACTGTACCACAGTATGCTTCTTCAAAGCAGTGGTTCCTTTACGC 720
 QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 DB 721 TGTATAGCTGCGGGGTCAACTGAACCTCAAGCGCCACAGACAGGATTTGTGGCGCGAG 780
 QY 261 SerAlaLeuProGlyValAlaTyrProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
 DB 781 AGCGCGCTCCCGGGGCTGGCCCTGGCAGGTGACCTGCGACGTCCAGAACGTCACGTG 840
 QY 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaHisCysValGluLys 300
 DB 841 TCGGAGGCTCCATCATCAACCCCGAGTGTGTCAGCGCCCACTCGTGGGAAAAA 900
 QY 301 ProLeuAsnAspProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320

DB 901 CCTCTTAACAATCCATGGCATTTGGACGGCATTTTGGGGGATTTTGAGACAATCTTTTCATG 960
 QY 321 PheTyrGlyValGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 DB 961 TTCTATGGAGCGGATCCCAAGTAGAAAAGTGAATTTCTCATCAATATGACTCCAAG 1020
 QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 DB 1021 ACCAAGAAACAATGACATTTGGCTGTATGAAGCTGCAGAAAGCTCTGACTTTTCAACGACCTA 1080
 QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
 DB 1081 GTGAACCAAGTGTGTGTGCCCAACCCAGGCAATGTCTGCGACGACAGACAGCTTCTGCTGG 1140
 QY 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaLeu 400
 DB 1141 ATTTCCGGGTGGGGGGGCGCCAGCGAGGAAAGGAAGACCTCAGAGTGTGTGAACGCTGCC 1200
 QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 DB 1201 AAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGAGATATGTCTATGACACACTGATC 1260
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 DB 1261 ACACGACGATGATCTGTGCGGCTTCTGCGAGGAACGTCGATTCCTGCCAGGTCGAC 1320
 QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpIleLeuIleGlyAspThrSer 460
 DB 1321 AGTGGAGGGCTCTGTGTCACCTTCGAGAAACAATATCTGGTGGCTGATAGGGGATCAAGC 1380
 QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 DB 1381 TGGGGTCTGTGCTGTGCCAAGCTTACAGACCAGGAGTGTACGGGAATGTGATGGTATTC 1440
 QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
 DB 1441 ACSGACTGGATTTATCGACAATGAGGGCAGACGGC 1476

RESULT 11

US-10-144-678A-931

; Sequence 931, Application US/10144678A

; Publication No. US20030157089A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Jiang, Yuqiu

; APPLICANT: Henderson, Robert A.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.

; APPLICANT: Day, Craig H.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Carter, Darrick

; APPLICANT: Li, Samuel X.

; APPLICANT: Wang, Aijun

; APPLICANT: Skeiky, Yasar A. W.

; APPLICANT: Hepler, William T.

; APPLICANT: Hural, John

; APPLICANT: McNeill, Patricia D.

; APPLICANT: Houghton, Raymond L.

; APPLICANT: Vinals y de Baesols, Carlota

; APPLICANT: Foy, Teresa M.

; APPLICANT: Watanabe, Yoshinhiro

; APPLICANT: Deng, Ta

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER

; FILE REFERENCE: 210121.427C28

; CURRENT APPLICATION NUMBER: US/10/144,678A

; CURRENT FILING DATE: 2002-08-12

; NUMBER OF SEQ ID NOS: 1033

SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 931
LENGTH: 1476
TYPE: DNA
ORGANISM: Homo sapiens
US-10-144-678A-931

Alignment Scores:
Pred. No.: 2,73e-306 Length: 1476
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 15 Gaps: 0

US-09-615-285B-2 (1-492) x US-10-144-678A-931 (1-1476)

```
QY 1 MetAlaLeuAenSerGlySerProAlaileGlyProTyrTyrGluAenHisGlyTyr 20
DB 1 ATGGCTTGAATCAGGCTCACCACAGCTATTGGACCTTACTATGAACACCTGATAC 60
QY 21 GlnProGluAenProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
DB 61 CAACCGGAAACCCCTATTCCGACAGCCCACTGTGTGCTCCCACTGTCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CCGGCTCAGTACTACCGCTCCCGCTGCGCCAGTACGCCCGCGGGTCTGTACGCGAGCT 180
QY 61 SerAenProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 181 TCAACCCCGCTGCTGACGAGCGCCCAATCCCATCCGAGGAGGTGCACCTCAAG 240
QY 81 ThrLysLysAlaLeuCysileThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
DB 241 ACTAAGAAAGACTGTGATCACCCTGACCTGGGGACCTTCTCGTGGGAGCTGGCTG 300
QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAenSerGlyLeuGlyCys 120
DB 301 GCGCTGGCCCTACTCTGGAAGTTATGGGAGCAGGTGCTCCCACTCTGGGATAGAGTGC 360
QY 121 AspSerSerGlyThrCysileAenProSerAenTrpCysAepGlyValSerHisCysPro 140
DB 361 GACTCCTCAGGTACCTGATCAACCCCTCTAATCTGGTGTGATGGGTGTGCACACTGCC 420
QY 141 GlyGlyGluAenGluAenArgCysValArgLeuTyrGlyProAenPheLeuGlnVal 160
DB 421 GCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAenSerTrpAenGluAenTyr 180
DB 481 TACTCATCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
QY 181 GlyArgAlaAlaCysArgAenMetGlyTyrLysAenAenPheTyrSerSerGlnGlyLe 200
DB 541 GGGCGGGCGGCTCGAGGAGCATGGGCTATAGAGATATATTTTACTCTAGCCAGGAGAT 600
QY 201 ValAenAenSerGlySerThrSerPheMetLysLeuAenThrSerAlaGlyAenValA 220
DB 601 GTGGATGACAGCGGATCCACAGCTTTATGAAATGAACCAAGTGCAGGCAATGTGCAT 660
QY 221 IleTyrLysLysLeuTyrHisSerAenAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 ATCTATAAAGAACTGTACACAGTGTGCTGTCTTCAAAAGCAGTGTGTCTTTTACGC 720
QY 241 CysIleAlaCysGlyValAenLysAenSerArgGlnSerArgLeValGlyGlyGlu 260
DB 721 TGTATAGCTGGGGGTCACTTGAATCAAGCGCCAGAGGAGGATGTGGGCGCGAG 780
QY 261 SerAlaLeuProGlyValAenTrpTyrGlnValSerLeuHisValGlnAenValHisVal 280
DB 781 AGCGGCTCCCGGGGGCTGACCTGGCAGGTGACCTGACCTGACAGACGCTCCACGCTG 840
```

```
QY 281 CysGlyGlySerIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
DB 841 TCGAGAGCTCCATCATCATCCCCGAGTGTGATCTGATCTGATCTGATCTGATCTGATCTG 900
QY 301 ProLeuAenAenProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
DB 901 CCTCTTAACAATCCATGGCATTTGGACCGCATTTGGCGGATTTGGAGACAAATCTTTCA 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGlyLysValIleSerHisProAenTyrAenSerLys 340
DB 961 TTCTATGAGCGGATACCAAGTAGAAAAGATTTCTCATCCAAATTTATGACTCCCAAG 1020
QY 341 ThrLysAenAenAenPheIleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAen 360
DB 1021 ACCAAGAACCAATGACATTTGGCTGATGAAGCTGCAAGAGCTCTGACTTTCAACGACCT 1080
QY 361 ValLysProValCysLeuProAenProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
DB 1081 GTGAACCAAGTGTCTGCCCCAACCCAGGATGATGTGTCAGCCAGAACAGCTCTGCTGG 1140
QY 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAenAlaAla 400
DB 1141 ATTTCCGGGTGGGGGGCCACCGAGGAGAAAGAAAGACCTCAGAAAGTGTGTAACGCTGC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAenSerArgTyrValTyrAenAenLeuIle 420
DB 1201 AAGGTGCTTCTCATTCAGACACAGAGATGCAACAGAGATATGTCTATGACAACTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAenValAenSerCysGlnGlyA 440
DB 1261 ACACGAGCATGATCTGTGCGGCTTCTGTCAGGGGAAAGCTGATTTGTCAGGGTGCAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAenAenIleTyrTrpLeuIleGlyAenThrSer 460
DB 1321 AGTGGAGGGCTCTGCTGCTCACTTCAAGAACAAATATCTGCTGCTGATAGGGGATACA 1380
QY 461 TrpGlySerCysAlaLysAlaTyrArgProGlyValTyrGlyAenValMetValPhe 480
DB 1381 TGGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
QY 481 ThrAenTrpIleTyrArgGlnMetArgAlaAenGly 492
DB 1441 ACGGACTGGATTTATCGACAAATGAGGGCAGACGGC 1476

RESULT 12
US-10-294-025-931
; Sequence 931, Application US/10294025
; Publication No. US20030185830A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C29
; CURRENT APPLICATION NUMBER: US/10/294,025
; CURRENT FILING DATE: 2002-11-12
; NUMBER OF SEQ ID NOS: 1038
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 931
; LENGTH: 1476
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-294-025-931

Alignment Scores:
Pred. No.: 2,73e-306 Length: 1476
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 15 Gaps: 0
```

US-09-615-285B-2 (1-492) x US-10-294-025-931 (1-1476)

```
QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 1 ATGGCTTTGACTCAGGCTCACCACAGCTATTGGACCTTACTATGAAAACCTGGATAC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 61 CAACCGGAAACCCCTATCCCGACAGCCCACTGTGGTCCCACTGTCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACTACCGCTCCCGCTGCCAGTACGCCCGGAGGCTCTCAGCAGGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCCGCTGTCTGACGAGGAGCCCAATCCCACTCCGAGCAGTGTGCACCTCAAAG 240
QY 81 ThrLysLeuAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyValAlaLeu 100
Db 241 ACTAAGAAAGCACTGTGCATCACTTGACCTGGGACCTTCTCTCGTGGGAGCTGGCTG 300
QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCGCTGGGCTACTCTCGAAGTTCATGGCAGCAAGTGTCTCACTCTGGGATAGATGC 360
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
Db 361 GACTCTCAGGTACCTGTGCATCAACCCCTCTAATCTGTGTGTGGCTGTCACTGGCCCC 420
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 421 GCGGGGAGGAGCAGAATCGGTGTGTGTGCTTACGGATCAAACTTCATCTTCAGGTG 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
Db 481 TACTCATCTCAGAGGAGTCTTGGCACCTCTGTGTCAGACGACTGGAACGAGAACTAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 541 GCGCGGGCGGCTCGCAGGACATGGCTATAGAAATAATTTTACTCTAGCAAGGAATA 600
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 601 GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACACAGTGGCGGCAATGTGAT 660
QY 221 IleTyrLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATAAAAACCTGTACCACAGTGTGCTGTCTTCAAAAGCAGTGGTTCCTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
Db 721 TGTATAGCTTGGGGGTCAACTTGAACCTCAAGCGCCAGACAGGAGTGTGGGGCGGAG 780
QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
Db 781 AGCGCGCTCCCGGGGCTGGCCCTGGGAGGTCAGCTGCACGCTCCAGAACGTCACAGTG 840
QY 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
Db 841 TCGGAGGCTCCATCATCACCCCGAGTGGATCGTGACAGCGGCCACTGCGTGGAAAAA 900
QY 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTTAACAATCCATGGCATTTGGACGCGATTTGGGGGATTTTGAGACAATCTTTTCATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 961 TTTCTATGAGCGCGATACCAAGTAGAAAAGTATTTCTCATCCCAATATTAGCTCCAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACATGATTCGCTGTATGAAGCTGACAGAGCCCTCTGACTTTCAACGACCTA 1080
```

```
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
Db 1081 GTGAACCAAGTGTCTGTCCCAACCCAGGCATGATGTGCAGCCAGAACAGCTCTGCTGG 1140
QY 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCGGGGTGGGGGCCACCGAGGAGAAAGGAAGACCTCAGAAAGTGTGAACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1201 AAGGTGTCTTCATTGAGACACAGAGATGCAACAGCAGATATGTCTATGCAACCTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACGAGCCATGATCTGTGCGGCTTCTGCGGGGACGTCGATTTTTCACAGGGTAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
Db 1321 AGTGGAGGCGCTCTGTCTCACTTCGAAGAACAAATATCTGTTGGCTGTATAGGGGATACAAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGTTTCTGCTGTGCGCAAGCTTACAGACCAGAGGTACGGGAATGTGATGTGTTTC 1440
QY 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
Db 1441 ACGGACTGGATTATTCGACAAATGAGGGCAGACGGC 1476
```

RESULT 13

US-09-759-143-930
; Sequence 930, Application US/09759143
; Patent No. US2002002248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09759,143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 930
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-759-143-930

Alignment Scores:
Pred. No.: 2,74e-306 Length: 1479
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: Gaps: 0

US-09-615-285B-2 (1-492) x US-09-759-143-930 (1-1479)

QY 1 MetAlaLeuAenSerGlySerProAlaIleGlyProTyrTyrGluAenHisGlyTyr 20
DB 1 ATGGCTTTTGAATCAGGCTCACCACAGCTATTGGACCTTACTATGAAACACCATGGATAC 60
QY 21 GlnProGluAenProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
DB 61 CACCGGGAACCCCTATCCCGCACACGCCACTGTGTGCCCACTGTCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CCGGCTCAGTACTACCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 180
QY 61 SerAenProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 181 TCCACCCCGTGTCTGTCAGCAGCAGCCAAATCCCATCCGGGACAGTGTGCACCTCAAG 240
QY 81 ThrLysValAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB 241 ACTAGAAAGCAGCTGTGCATCCTTGACCTCGGGACCTTCTCTGTGGAGCTGGCTG 300
QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAenSerGlyIleGluCys 120
DB 301 GCCGTGCGCTACTCTGAAAGTTCATGGGAGCAGAGTGTCTCCAACTCTGGGATAGAGTGC 360
QY 121 AspSerSerGlyThrCysIleAenProSerAenTyrCysAspGlyValSerHisCysPro 140
DB 361 GACTCCTCAGTACTCTGCATCAACCCCTTAACTGGTGTGTGAGTGTGCACACTGCCCC 420
QY 141 GlyGlyGluAenAenArgCysValArgLeuTyrGlyProAenPheIleLeuGlnVal 160
DB 421 GCGGGGAGGAGCAGAAATCGGTGTGTTCGCTCTACGGATCAAACTTCATCTTCAGGTG 480
QY 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTrpAenGluAenTyr 180
DB 481 TACTCATCTCAGAGAGAGTCTGGCACCTGTGTGCCAGACACACTGGAAACGAGACTAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAenAenPheTyrSerGlnGlyIle 200
DB 541 GCGCGGGCGGCTCGCAGGACATGGGCTATAAGATAAATTTTACTCTAGCCCAAGGAATA 600
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAenThrSerAlaGlnValAsnValAsp 220
DB 601 GTGGATGACAGCGATCCACAGCTTTTGAACACTGACACACAGTCCCGGCGATGTGCAT 660
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 ATCTATAAAAACTGTACACAGTGTGCTGTCTTCAAAAGCAGTGTGTCTTTTACGC 720
QY 241 CysIleAlaCysGlyValAenAenSerSerArgGlnSerArgIleValGlyGlyGlu 260
DB 721 TGTATAGCTCGGGGTCACTTGAACCTCAAGCGCCGACAGCAGGATTTGGGCGGCGAG 780
QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAenValHisVal 280
DB 781 AGCGCGCTCCCGGGGCTGGCTCGCAGAGTGTGCTGTCTTCAAAAGCAGTGTGTCTTTACGC 840
QY 281 CysGlyGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
DB 841 TGGGAGGCTCCATCATCACCCCGAGTGTGATGTGACAGCGCCGCTGCTGGGAAAAA 900
QY 301 ProLeuAenAenProTyrHisThrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
DB 901 CCTCTTAACAATCATGCGATTTGACGCGCATTTTGGGGGATTTTGAGACAACTTTTCATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAenTyrAspSerLys 340
DB 961 TTCTATGGAGCGGATACCAAGTAGAAAAAGTGTATCTCATCCAAATTTATGACTCCAAAG 1020
QY 341 ThrLysAenAenAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAspLeu 360
DB 1021 ACCAAGAACAAATGATGCTGCTGATGAAGCTGCAAGAGCTCTGTGACTTTTCAACGACCTA 1080

QY 361 ValLysProValCysLeuProAenProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
DB 1081 GTGAACACAGTGTGTCTGCCCAACCCAGGATGATGTCTGCAGCAGAACAGCTCTGCTGG 1140
QY 381 IleSerGlyTyrGlyAlaThrGluLysGlyLysThrSerGluValValLeuAenAlaAla 400
DB 1141 ATTTCGGGTGGGGGGCCACCGAGGAGAAAGGAGACCTCAGAAGTGTCTGAACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAenSerArgTyrValTyrAspAenLeuIle 420
DB 1201 AGGTGCTTCTCATTTAGACACAGAGATGCAACAGCAGATATGTCTATGACACCTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAenValAspSerCysGlnGlyAsp 440
DB 1261 ACACACAGCATCATCTGTGCCGCTCTCTGCAGGAGAAACGTCGATTTCTTGCAGGGTAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAenAenIleTyrTrpLeuIleGlyAspThrSer 460
DB 1321 AGTGGAGGCGCTCTGGTCACCTTCGAGAGAACATATCTGGTGGCTGATAGGGGATACAAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAenValMetValPhe 480
DB 1381 TGGGGTCTTGGGTGTGCCAAAGCTTTACAGACAGGAGTGTACGGGAAATGTGTATGGTATTC 1440
QY 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
DB 1441 ACGGACTGGATTATCGACAAATGAGGGCAGACGGC 1476

RESULT 14

US-09-780-669-930
; Sequence 930, Application US/09780669
; Patent No. US20020051977A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Koughton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C24
; CURRENT APPLICATION NUMBER: US/09/780,669
; CURRENT FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 943
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 930
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-780-669-930

Alignment Scores:
Pred. No.: 2,74e-306 Length: 1479
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: Gaps: 0

US-09-615-285B-2 (1-492) x US-09-780-669-930 (1-1479)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 1 ATGGCTTTGAGCTCAGGCTCAGCCAGCTATTGGACCTTACTATGAAACCATGGATAC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
DB 61 CAACCGGAAACCCCTATCCCGACAGCCCAATCCCAATCCCGACAGTGTGAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CGGGCTCAGTACCTACCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 181 TCCAAACCCCGCTGTCTGACGAGCAGCCCAATCCCAATCCCGACAGTGTGATCAAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB 241 ACTAAGAAAGCACTGTGATCATCCTTGACCTGGGGACCTTCTCTGGGGAGCTGGCGCTG 300
QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyLysGluCys 120
DB 301 GCGCTGGCTACTCTGGAGTTCATGGCAGCAGTGTCTCAACTCTGGGATAGAGTGC 360
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnThrProCysAspGlyValSerHisCysPro 140
DB 361 GACTCCTCAGGTACCTGTCATCAACCCCTTCACTGGTGTGATGGCGTGTCACTGCCCC 420
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
DB 421 GCGCGGAGAGCAGATCCGCTGTGTTCCTCTACGATCAAACTTCATCTCTCAGGTG 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspTrpAsnGluAsnTyr 180
DB 481 TACTCATCTCAGAGGAAGTCTTGGACCTGTGTGCCAAGCAGCAGTGGAAACGAACTAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
DB 541 GGGCGGGCGCTCGAGGACATGGCTATPAGATATATTTTACTCTAGCCAAAGGATA 600
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB 601 GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACACAAAGTGGCGGCAATGTGAT 660
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 ATCTATAAATACTGTACACAGTATGCTTCTTCAAAAGCAGTGGTTCTTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
DB 721 TGTATAGCTCGCGGCTCAACTTGAACCTCAAGCGCCAGCAGCAGGATTTGGCGGCGAG 780
QY 261 SerAlaLeuProGlyAlaTyrProGlnValSerLeuHisValGlnAsnValHisVal 280
DB 781 ACGGGCTCCCGGGGCTGGCCCTTGGCAGTCACTGCTGACGTCGACGTCGACGTCGACG 840
QY 281 CysGlyGlySerIleThrProGlnTrpIleValThrAlaAlaHisCysValGluLys 300
DB 841 TGGCGAGGCTCCATCATCCCGCTGGATCGTGTGACAGCGCGCCCACTGCTGTGAAAAA 900
QY 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
DB 901 CTTCTTAACATCCATGGCATTTGGAGCATTTGGGGGATTTTGGAGCATCTTTTCATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
DB 961 TTCTATGGAGCGGATACCAAGTAGAAAAAGTATTTCTCATCCAAATTTATGATCCAAAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360

DB 1021 ACCAAGAACATGATTCGCTGATGAAGCTGCAGAGCCTCTGACTTTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
DB 1081 GTGAACACCATGTGTCTGCCAACCCAGGCAATGATGTGCAGCCAGAACAGCTCTGTGG 1140
QY 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
DB 1141 ATTTCCGGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAAAGTGTGAAAGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
DB 1201 AAGGTGCTTCTCATGACACACAGAGATGCAACAGCAGATATGTCTATGACAACTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
DB 1261 ACACGAGCCATGATCTGTGCGGCTTCTGCGAGGGAACGTCGATTTCTTCCAGGTTGAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
DB 1321 AGTGGAGGCGCTCTGGTCACCTTCGAAGAACAAATATCTGGTGGCTGATAGGGGATACAAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
DB 1381 TGGGGTTCGTGCTGCCAACAGCTTACAGCACGAGGTGTACGGGAATGTGATGTTATTC 1440
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
DB 1441 ACGGACTGATTTATCGACAAATGAGGCGACAGCGC 1476

RESULT 15

US-09-822-827-930
; Sequence 930 Application US/09822827
; Patent No. US20020081680A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C1
; CURRENT APPLICATION NUMBER: US/09/822,827
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 930
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-827-930

Alignment Scores:
Pred. No.: 2,74e-306 Length: 1479
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 9 Gaps: 0

US-09-615-285B-2 (1-492) x US-09-822-827-930 (1-1479)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 1 ATGGCTTTGAACTCAGGCTCAGCCAGCTATTGGACCTTACTATGAAACCATGGATAC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
DB 61 CAACCGGAAACCCCTATCCCGACAGCCCAATCCCAATCCCGACAGTGTGATCAAAG 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CGGGCTCAGTACTACCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80

Db 181 TCCAAACCCGCTGCTGTCACGAGCCCAATCCCATCCGGGACAGTGTGCACCTCAAAG 240
Qy 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 241 ACTAAGAAGCACTGTGCATCACTTGACCTGGGGACCTTCCTCGTGGAGGTGCGCTG 300
Qy 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCGGTGGGCTACTCTGGAAGTTCATGGGCGAGAGTCTCCAACTCTGGGATAGAGTGC 360
Qy 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
Db 361 GACTCTCAGGTACCTGCATCAACCCCTCTAACTGGTGTGATGGCGTGTCACTGCCCC 420
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 421 GCGGGGAGGAGCGAAGTGGTGTGCTGCTCTACGGATCAAACTTCATCCTTCAGGTG 480
Qy 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
Db 481 TACTCATCTCAGAGGAGTCTGGCACCTCTGTGTGCCAAGACGACTGGACAGGAACCTAC 540
Qy 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 541 GCGCGGGCGGCTCTCAGGAGACATGGGCTATAAGAAATAATTTTACTCTAGCCCAAGGAATA 600
Qy 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 601 GTGGATGACAGCGGATCCACGAGCTTTATGAACTGACACCAAGTGGCGGCAATGTGAT 660
Qy 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATAAAAACTGTACACAGATGATGCTGTCTTCAAAGCAGTGGTTCCTTTACGC 720
Qy 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
Db 721 TGTATAGCTCGGGGGTCACTGAACCTCAAGCCGCCAGACGAGATTGTGGCGCGAG 780
Qy 261 SerAlaLeuProGlyAlaTyrProTyrPheValSerLysHisValGlnAsnValHisVal 280
Db 781 AGCGGCTCCCCGGGGGCTGGCCCTGGCAGTCAAGCTGACGCGGCCACTCGGTGGAAAA 900
Qy 281 CysGlyGlySerIleIleProGluTyrIleValThrAlaAlaHisCysValGluLys 300
Db 841 TCGGGAGGCTCCATCATCAACCCCGAGTGGATCTGACAGCGGCCACTCGGTGGAAAA 960
Qy 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTTAAACAATCCATGGCATGGACGGCATTTGGGGGATTTTGGACAAATCTTTTCATG 960
Qy 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 961 TTCTATGGAGCGGATACCAAGTAGAABAAGTATTTCTCATCCAAATTTAGACTCCAAG 1020
Qy 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACAATGACATTTGCGCTGATGAAGTGCAGAGGCTCTGACTTTCAACGACCTA 1080
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
Db 1081 GTGAACCCAGTGTGTCTGCCCAACCCAGGCATGATGTGTCAGCCAGACAGCTCTGCTGG 1140
Qy 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCCGGGTGGGGGCGCCAGAGAGAGGAGGAGACCTCAGAAAGTGTGAACGCTGCC 1200
Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1201 AAGGTGCTTCTCATTTGACACACAGAGATGCAACAGCAGATATGTCTATGACAACTGATC 1260
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACGAGCCATGATCTGTGCGGCTTCTGTGAGGGGAAACGTCGATTTCTGCCAGGGTGAC 1320

Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
Db 1321 AGTGGAGGGCTCTGGTCACTTCGAGAAACAATATCTGGTGGCTGATAGGGATACAAGC 1380
Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGGTTCTGGCTGTGCCAAAGCTTACAGACCAGGAGTGTACGGGAATGTGATGGTATTC 1440
Qy 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
Db 1441 ACGGACTGGATTTATCGACAAATGAGGGGACAGCGC 1476

Search completed: June 1, 2004, 19:17:39
Job time : 749 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 1, 2004, 14:46:00 ; Search time 131 Seconds

(without alignments)
2084.243 Million cell updates/sec

Title: US-09-615-285B-2

Perfect score: 2717

Sequence: 1 MALNSGPPAIGPYENHGY.....VYGNVWFTDIYRQMRADG 492

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=xlh
-Q=/cgn2_1/USPTO.spool/US09615285/runat_01062004_132134_19976/app_query.fasta_1.647
-DB=issued Patents NA -QFMT=fastat -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09615285.cgn2_1.69 @runat_01062004_132134_19976 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEFTIMEOUT=120 -WARN_TIMEOUT=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/ECTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2717	100.0	1479	US-09-342-749-1	Sequence 1, Appli
2	2717	100.0	1479	US-09-691-840-1	Sequence 1, Appli
3	2696	99.2	2479	US-09-342-749-29	Sequence 29, Appl
4	2696	99.2	2479	US-09-691-840-29	Sequence 29, Appl
5	2696	99.2	2479	US-09-685-166A-894	Sequence 894, App
6	1569	57.7	1077	US-08-807-151-2	Sequence 2, Appli
7	1569	57.7	1077	US-09-478-957-2	Sequence 2, Appli
8	1484	54.6	1001	US-09-016-434-963	Sequence 963, App
9	1165	42.9	683	US-09-685-166A-896	Sequence 896, App
10	888	32.7	2413	US-09-518-046-1	Sequence 1, Appli
11	863.5	31.8	2544	US-09-518-046-3	Sequence 3, Appli
12	858	31.6	2416	US-09-261-416-1	Sequence 1, Appli

13	687.5	25.3	2079	4	US-09-656-002-1	Sequence 1, Appli
14	687.5	25.3	2079	4	US-09-851-588-5	Sequence 5, Appli
15	2038	24.9	2038	3	US-09-008-271A-18	Sequence 18, Appli
16	676.5	24.9	2081	4	US-09-851-588-7	Sequence 7, Appli
17	660	24.3	2581	1	US-08-200-900A-1	Sequence 1, Appli
18	660	24.3	2581	5	PCT-US94-00616-1	Sequence 1, Appli
19	590.5	21.7	1739	2	US-08-681-151-2	Sequence 2, Appli
20	585.5	21.5	2363	4	US-09-742-703-3	Sequence 3, Appli
21	580	21.3	1783	3	US-09-510-738A-188	Sequence 188, App
22	580	21.3	1783	4	US-09-861-966-188	Sequence 188, App
23	558.5	20.6	3147	2	US-09-027-337-1	Sequence 1, Appli
24	558.5	20.6	3147	4	US-09-644-600-1	Sequence 1, Appli
25	558.5	20.6	3147	4	US-09-644-600-18	Sequence 18, Appli
26	558.5	20.6	3147	4	US-09-654-600A-1	Sequence 1, Appli
27	558.5	20.6	3147	4	US-09-654-600A-18	Sequence 18, Appli
28	557	20.4	1515	4	US-09-820-002-1	Sequence 1, Appli
29	554.5	20.4	2152	4	US-09-023-655-157	Sequence 157, App
30	540.5	19.9	1605	2	US-09-008-846-1	Sequence 1, Appli
31	527.5	19.4	1460	4	US-09-370-838-80	Sequence 80, Appli
32	527.5	19.4	1462	4	US-09-370-838-55	Sequence 55, Appli
33	527.5	19.4	1517	1	US-08-508-448C-15	Sequence 15, Appli
34	527.5	19.4	2790	4	US-09-370-838-79	Sequence 79, Appli
35	513.5	18.9	696	1	US-08-508-448C-24	Sequence 24, Appli
36	500.5	18.4	959	4	US-09-023-942A-25	Sequence 25, Appli
37	497.5	18.3	901	1	US-08-508-448C-9	Sequence 9, Appli
38	497.5	18.3	2900	2	US-09-027-337-9	Sequence 9, Appli
39	497.5	18.3	2900	4	US-09-644-600-9	Sequence 9, Appli
40	497.5	18.3	2900	4	US-09-654-600A-9	Sequence 9, Appli
41	490	18.0	1553	4	US-09-280-116-10	Sequence 10, Appli
42	485.5	17.9	933	4	US-09-023-942A-29	Sequence 29, Appli
43	475	17.5	1225	4	US-09-734-675-1	Sequence 1, Appli
44	471	17.3	1110	4	US-09-386-653A-1	Sequence 1, Appli
45	468.5	17.2	980	4	US-09-023-942A-30	Sequence 30, Appli

ALIGNMENTS

RESULT 1

US-09-342-749-1
; Sequence 1, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMPRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1476)
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (724)
; OTHER INFORMATION: Listed as T in GenBank Accession NO. U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (985)
; OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (1347)
; OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329

```

FEATURE:
NAME/KEY: conflict
LOCATION: (1466)
OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1471)
OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329.
FEATURE:
NAME/KEY: allele
LOCATION: (478)
OTHER INFORMATION: This base can be G or A with G being the more
OTHER INFORMATION: common allele. The codon will change from Val to
OTHER INFORMATION: Met.
FEATURE:
NAME/KEY: allele
LOCATION: (777)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. The codon is unaffected with both
OTHER INFORMATION: alleles encoding Gly.
FEATURE:
NAME/KEY: allele
LOCATION: (768)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
FEATURE:
NAME/KEY: allele
LOCATION: (834)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
FEATURE:
NAME/KEY: allele
LOCATION: (625)
OTHER INFORMATION: This base can be T or A with T being the more
OTHER INFORMATION: common allele. The codon will change from Phe to
OTHER INFORMATION: Ile
US-09-342-749-1

Alignment Scores:
Pred. No.: 9,21e-258 Length: 1479
Score: 2717.00 Matches: 492
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-615-285B-2 (1-492) x US-09-342-749-1 (1-1479)
Qy 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 1 ATGGCTTTGAACTCAGGGTCAACCAGGCTATTGGACCTTACTATGAAAACCATGGATAC 60
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 61 CAACCGGAAACCCCTATCCCGCACAGCCCACTGTGTCCTCCACCTGTCTACAGGTGCAT 120
Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACTACCCGTCCTCCCGTCCAGTACGCCCGGAGGCTCTGACGCGAGCT 180
Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyTyrValCysThrSerLys 80
Db 181 TCCAAACCCGCTGCTCAGCGAGCCCAATCCCATCCGACAGAGTGGACCTTCAAG 240
Qy 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 241 ACTAAGAAGACCTGTGCATCAGCTTGAACCTGGGAGCCCTTCTCTGTGGAGCTGCGCTG 300
Qy 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCCGCTGCCCTACTCTGGAGTTCATGGCAGCAGCAAGTGTCCAACTCTGGGATAGATGC 360
Qy 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140

```

```

Db 361 GACTCTCTCAGTACCTGTCATCAACCCCTCTAACTGGTGTGATGGCTGCACACTGCCCC 420
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 421 GCGCGGGAGGACGAGAAATCGGTGTGTTCGCTCTACGACCAAACTTCATCTCTCAGGTG 480
Qy 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Db 481 TACTCTCTCAGAGGAGTCTGTCACCTGTGTGCCAGACGCTGGACAGAACTAC 540
Qy 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 541 GCGCGGGGCGCTCGCAGGACATGGGCTATAGAATAATTTTTTACTCTAGCCAGAAATA 600
Qy 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 601 GTGGATGACAGCGGATCCACCGCTTTATGAACCTGAACACAGTGGCGGCAATGTCGAT 660
Qy 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATAAAAACTGTACCAAGTGTCTTCTCAAAAGCAGTGGTTCCTTTACGC 720
Qy 241 CysIleAlaCysGlyWalAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 721 TGTATAGCTCGGGGTCACTTGAACTCAAGCGCCAGACAGGATCGTGGCGCGAG 780
Qy 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 781 AGCGCGCTCCCGGGGCTGCGCTGGCAGGTGAGCTGCACGTGCAGAACGTCCACGTG 840
Qy 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db 841 TCGGAGGCTCCATCATCATCCCCCGAGTGGATCGTGACAGCGCCCACTGCGTGGAAAA 900
Qy 301 ProLeuAsnAsnProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTTAACAATCCATGGCATTTGGACGGCAATTTTCGGGGATTTTGACAAATCTTTCATG 360
Qy 321 PheTyrGlyAlaGlyTyrGlnValGlyValIleSerHisProAsnTyrAspSerLys 340
Db 961 TTCTATGGGCGGATACCAAGTAGAAAAAGTATTCTCATCCAAATATGACTCCAAG 1020
Qy 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACAATGACATTTGCTGATGAGCTCGAGAGCTCTGACTTCCAGACCTA 1080
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1081 GTGAAACCAAGTGTCTGCGCCACACCCAGGATGATGCTGCAGCCAGACAGCTCTGCTGG 1140
Qy 381 IleSerGlyTyrGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCCGGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAGTGTGACGCTGCC 1200
Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1201 AAGGTGCTTCTCAITGAGACACAGAGATGCAACAGCAGATATGCTATGACACCTGATC 1260
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACAGCAGCATGATCTGTGCGGCTTCTCGAGGGAGACGTCGATTTCTTTCAGGCTGAC 1320
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1321 AGTGAGGGCTCTGGTCACTTCGAGAACAAATATCTGTGGCTGATAGGGATACAGC 1380
Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGGTCTGGCTGTGCAAAAGCTTACAGACAGGAGTGTACGGGAATGTGATGTTATTC 1440
Qy 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

```

Db 1441 ACGACTGGATTATCGACAATAGGGCGACAGCGC 1476

RESULT 2

US-09-691-840-1
 ; Sequence 1, Application US/09691840
 ; Patent No. 6444419
 ; GENERAL INFORMATION:
 ; APPLICANT: Wong, Alexander K.C.
 ; APPLICANT: Tavtighian, Sean V.
 ; APPLICANT: Teng, David H.-F.
 ; APPLICANT: Myriad Genetics, Inc.
 ; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
 ; FILE OF INVENTION: 2318-202
 ; CURRENT APPLICATION NUMBER: US/09/691,840
 ; CURRENT FILING DATE: 2000-10-18
 ; PRIOR APPLICATION NUMBER: US/09/342,749
 ; PRIOR FILING DATE: 1999-06-29
 ; PRIOR APPLICATION NUMBER: US 60/091,044
 ; PRIOR FILING DATE: 1998-06-29
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1479
 ; TYPE: DNA

ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)..(1476)
 NAME/KEY: conflict
 LOCATION: (724)
 OTHER INFORMATION: Listed as T in GenBank Accession NO. U75329
 NAME/KEY: conflict
 LOCATION: (985)
 OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
 NAME/KEY: conflict
 LOCATION: (1347)
 OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
 NAME/KEY: conflict
 LOCATION: (1466)
 OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329
 NAME/KEY: conflict
 LOCATION: (1471)
 OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329
 NAME/KEY: allele
 LOCATION: (478)
 OTHER INFORMATION: This base can be G or A with G being the more common allele. The codon will change from Val to Met.
 NAME/KEY: allele
 LOCATION: (777)
 OTHER INFORMATION: This base can be C or T with C being the more common allele. The codon is unaffected with both alleles encoding Gly.
 NAME/KEY: allele
 LOCATION: (768)
 OTHER INFORMATION: This base can be C or T with C being the more common allele. This is a silent polymorphism.
 NAME/KEY: allele
 LOCATION: (834)
 OTHER INFORMATION: This base can be C or T with C being the more common allele. This is a silent polymorphism.
 NAME/KEY: allele
 LOCATION: (625)
 OTHER INFORMATION: This base can be T or A with T being the more common allele. The codon will change from Phe to Ile
 US-09-691-840-1

Alignment Scores:
 Pred. No.: 9 216-258 Length: 1479
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0

DB: 4 Gaps: 0

US-09-615-285B-2 (1-492) x US-09-691-840-1 (1-1479)

QY	1	MetAlaLeuAenSerGlySerProAlaIleGlyProTyTyTyGluAenHisGlyTyT	20
DB	1	ATGGCTTTGAACCTCAGGGTCAACCCAGCTATTGGACCTTACTATGAAACCATGGATC	60
QY	21	GlnProGluAenProTyTyProAlaGlnProThrValValProThrValTyGluValHis	40
DB	61	CAACCGGAAACCCCTATCCCGCACAGCCACCTGTGTGCTCCCACTGTCTACGAGGTGCAT	120
QY	41	ProAlaGlnTyTyTyProSerProValProGlnTyTyAlaProArgValLeuThrGlnAla	60
DB	121	CGGGCTCAGTACTACCGTCCCGTCCCGCAGTACGCCCGGAGGTCCTGACGCGAGCT	180
QY	61	SerAenProValValCysThrGlnProIysSerProSerGlyThrValCysThrSerIys	80
DB	181	TCCAAACCCCGTGTGTGCAGCAGCCCAATCCCATCCCGGACAGTGTGCACCTCAAG	240
QY	81	ThrIysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu	100
DB	241	ACTAAGAAAGCACTGTGCATCACCTTGACCTGGGGACCTTCCTCGTGGAGCTGGCTG	300
QY	101	AlaAlaGlyLeuLeuTyTyPheMetGlySerIysCysSerAenSerGlyIleGluCys	120
DB	301	GGCGTGGCTTACTGTGAAGTTTCATGGCAGCAAGTGTCTCCAACTCTGGATAGAGTGC	360
QY	121	AppSerSerGlyThrCysIleAenProSerAenTyTyCysAenGlyValSerHisCysPro	140
DB	361	GACTCCTCAGGTACTGTGCATCAACCTCTAACTGTGTGTGATGGCTGTGCACACTGCCCT	420
QY	141	GlyGlyGluAenGluAenArgCysValArgLeuTyTyGlyProAenPheIleLeuGlnVal	160
DB	421	GGCGGGGAGGAGGAGATCGGTGTGTCTTACGGACCAACTTCATCTTCAGGTG	480
QY	161	TyrSerSerGlnArgIysSerTrpHisProValCysGlnAenSerTrpAenGluAenTyT	180
DB	481	TACTCATCTCAGAGAAAGTCTGTGCACCTCTGTGTGCAAGACGACTGGAGACGAGACTAC	540
QY	181	GlyArgAlaAlaCysArgAenMetGlyTyTyLysAenAenPheTyTySerSerGlnGlyIle	200
DB	541	GGCGGGGGGGCTCAGGGACATGGGCTATAGAAATATTTTACTCTAGACCAAGGATA	600
QY	201	ValAenSerSerGlySerThrSerPheMetLysLeuAenThrSerAlaGlyAenValAen	220
DB	601	GTGGATCAGACGGATCCACGAGCTTTATGAAACTGAACACAGTCCCGGCAATGTTCGAT	660
QY	221	IleTyTyLysLysLeuTyTyHisSerAspAlaCysSerSerLysAlaValValSerLeuArg	240
DB	661	ATCTATAAAACCTGTACACAGTGTGTCTTCAAGAGCAGTGTCTTTTACGC	720
QY	241	CysIleAlaCysGlyValAenLeuAenSerArgGlnSerArgIleValGlyGlyGlu	260
DB	721	TGTATAGCTTGGGGGTCAACTTGAACCTCAAGCCCGCAGCAGGATCTGGGGCGGAG	780
QY	261	SerAlaLeuProGlyAlaTyTyProTrpGlnValSerLeuHisValGlnAenValHisVal	280
DB	781	AGCGGCTCCCGGGGGCTGTGCCTGTGCAGTGTGCAGTCCAGAACCTCCACGCTG	840
QY	281	CysGlyGlySerIleIleThrProGluTyTyIleValThrAlaAlaHisCysValGluIys	300
DB	841	TCCGGAGCTCCATCATCACCCCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	900
QY	301	ProLeuAenAenProTrpHisTrpThrAlaPheAlaGlyIleLeuAenGlnSerPheMet	320
DB	901	CCTCTTAACATCCATGGCATTTGGAGCGGATTTTGGGGGATTTTGGAGCAATCTTCATG	960
QY	321	PheTyTyGlyAlaGlyTyTyGlnValGluIysValIleSerHisProAenTyTySerIys	340
DB	961	TTCTATGGAGCCGATACCAAGTAGAAAGTAGATTTCTATCCCAATTTATGACTCCAG	1020
QY	341	ThrLysAenAenAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAspLeu	360

```

Db 1021 ACCAGACACATGACATGGCTGATGAGCTGCAGAGCCTCTGACTTTCAAGACCTA 1080
QY 361 VallysProValCysLeuProAsnProGlyMetMetLeuGlnProGluLeuCytrp 380
Db 1081 GTGAACACAGTGTCTGCGCCACCCAGGCGATGATGCTGCAGCCAGACAGCTCTGTCTGG 1140
QY 381 IleserGlyTrpGlyAlaThrGluGluCysGlyLysThrSerGluValLeuAsnAlaAa 400
Db 1141 ATTTCCGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAAAGTCTGAACGCTGCC 1200
QY 401 LysValLeuLeuLeuLeuThrGlnArgCysAsnSerArgTyrValTyrAsnLeuLeu 420
Db 1201 AAGTGTCTTCTCATTTGAGACACAGATGACACAGCAGATATGCTATGACACCTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACAGCCTATGATCTGTGCGGCTTCTGCGAGGAGACCTGCTGATTTCTGCCAGGGTAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnLeuLeuTrpTrpLeuLeuGlyAspThrSer 460
Db 1321 AGTGGAGGGCTCTGGTCACTTCCAAAGAACATATCTGTGGCTGATAGGGGATACAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGGTCTGGCTGTGCCAAGCTTACAGACAGGAGTGTACGGGAATGTGATGCTATTC 1440
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1441 ACGGACTGGATTATCGACAAATGAGGGGACAGCGC 1476

RESULT 3
US-09-342-749-29
; Sequence 29, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtighian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-342-749-29

Alignment Scores:
Pred. No.: 2,42e-255 Length: 2479
Score: 2696.00 Matches: 486
Percent Similarity: 99.80% Conservative: 5
Best Local Similarity: 98.78% Mismatches: 1
Query Match: 99.23% Indels: 0
DB: 3 Gaps: 0

US-09-615-285b-2 (1-492) x US-09-342-749-29 (1-2479)
QY 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 57 ATGGCTTGAACCTCAGGCTCACCACAGCTATTTGACCTTACTATGAAACCATGATATC 116
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 117 CAACCGGAACCCCTATCCGCCACAGCCCACTGTGTCCCACTGTCTACGAGGTGCAT 176
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60

```

```

177 CCGGCTCAGTACTACCGTCCCCCGTGCAGCCAGTACGCCCGAGGGTCTCTGCGAGGCT 236
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 237 TCCAAACCGGCTGCTGTCAGCGAGCCCAATCCCATCCCGGACAGTGTGCACTCAAG 296
QY 81 ThrLysValAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 297 ACTAAGAAAGACACTGTGTCATCACTTGACCTTGGGACCTTCTCTGTTGGAGCTCGCTG 356
QY 101 AlaIaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 357 GCGCTGCGCTACTCTTGAAGTTTCATGGGACGCAAGTGTCTCCAACTCTGGATAGAGTGC 416
QY 121 AspSerSerGlyThrCysIleLeuAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
Db 417 GACTCCTCAGTACCTGCATCAACCCCTCTAACTGGTGTGATGGGTGTGCACACTGCCCC 476
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 477 GCGGGGAGGACGAGAAATCGGTGTCTGCGCTCTACGACCAAACTTCTATCTTCAGATG 536
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Db 537 TACTCATCTCAGAGGAAGTCTCTGGCACCTGTGTGCCAAGACGACTGGAAACGAGAACTAC 596
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 597 GCGCGGGCGGCTTCAGGACATGGGCTATAGAAATATTTTTACTCTTACGCAAGAAATA 656
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 657 GTGGATGACAGCGGATCCACAGCTTTATGAACATGAACACAAAGTCCGCGCAATGTGCAT 716
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 717 ATCTATAAAAAACTGTACACAGTGTGCTTCTTCCAAAGCAGTGTCTTCTTACGC 776
QY 241 CysAlaAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
Db 777 TGTTTAGCTTGGGGGTCACTTGAACCTCAAGCCCGCAGACAGATCTGTGGGGGTGAG 836
QY 261 SerAlaLeuProGlyAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
Db 837 AGCGCGCTCCGCGGGGCTTGGCCCTGGCAGGTCAGCTGCACGTCAGAACGTCACAGTG 896
QY 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db 897 TGGGAGGGCTCATCATCACCCCGAGTGGATCGTGACAGCCGCCCACTGCGTGGAAAAA 956
QY 301 ProLeuAsnAsnProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 957 CTTCTTAAACAATCCATGGCATTTGGACGGCATTTGCGGGATTTTGAGACAATCTTTCATG 1016
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 1017 TTCTATGAGGCGGATACCAAGTACAAAAGTATTTCTCATCAATATATGATCTCCAG 1076
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1077 ACCAAGAACAATGACATTTGCGCTGATGAAGTGCAGAAAGCTCTGACTTTCAACGACCTA 1136
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1137 GTGAACACAGTGTGTCTGCCCAACCCAGGATGATGTGTCAGCCAGAACAGCTCTGCTGG 1196
QY 381 IleserGlyTrpGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAa 400
Db 1197 ATTTCCGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAAAGTGTGCTAAGCGTGC 1256
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420

```

Db 1257 AAGGTGCTTCTCATTGACACACAGAGATGCAACAGCAGATATGTCTATGCAACACCTGATC 1316
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyValAspSerCysGlnGlyValAsp 440
Db 1317 ACACCGACCATGATCTGTGCGGGCTTCTGAGGGGACGTCGATTTCTTGCCAGGGGTGAC 1376
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1377 AGTGAGGGGCTCTGGTCACATCGAACCAACAATATCTGGTGGCTGATAGGGGATACAAGC 1436
Qy 461 TrpGlySerGlyCysAlaGlyAlaValArgProGlyValTrpGlyValMetValPhe 480
Db 1437 TGGGGTCTGGCTGTGCGCAAGCTTACAGACAGAGGTACGGGAATGTGATGCTATTC 1496
Qy 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1497 ACGGACTGATTTATCGACAAATGAAGCAACAGGC 1532

RESULT 4
US-09-691-840-29
; Sequence 29, Application US/09691840
; Patent No. 6444419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TWPRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/691,840
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342,749
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091,044
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-691-840-29

Alignment Scores:
Pred. No.: 2,428-255 Length: 2479
Score: 2696.00 Matches: 486
Percent Similarity: 99.80% Conservative: 5
Best Local Similarity: 98.78% Mismatches: 1
Query Match: 99.23% Indels: 0
DB: 4 Gaps: 0

US-09-615-285B-2 (1-492) x US-09-691-840-29 (1-2479)

Qy 1 MetaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 57 ATGGCTTTGAATCAGGGGTCAACAGGATATGGACCTTACTATGAAACCATGGATAC 116
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
Db 117 CAACCGGAAACCCCTATCCCGCAGACGCCACTGTGGTCCCACTGTCTACAGGTGCAT 176
Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 177 CCGGCTCAGTACTACCGCTCCCGCTGCGCCAGTAGTACGCCCGGAGGCTCTGACGCGGCT 236
Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 237 TCCAAACCCGCTCGTTCGACGAGCGGCGCAATCCCATCCGAGACAGTGTGACCTCAAAG 296
Qy 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 297 ACTAGAAAGCACTGTGCGATACCTTGACCTGGGACCTTCTCTGTTGGAGCTGCGGTG 356

Qy 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 357 GCGCTGGGCTACTCTGGAAAGTTCTGGGACAGCAAGTGTCTCCAACTCTCTGGGATAGAGTGC 416
Qy 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
Db 417 GACTCTCTCAGGTACCTGTCATCAACCCCTCTAATCTGGTGTGATGGCGTGTCACTGCCCC 476
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 477 GCGGGGAGGACGAGAATCGGTGTGTTCGCTCTACGGACCAAACTTCTATCTCTCAGATG 536
Qy 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Db 537 TACTCATCTCAGAGGAAGTCTGGCACCTCTGTGCGCAAGACACTGGAAACGAGAACTAC 596
Qy 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 597 GCGCGGGGCGCTGCAGGCACATGGCTATAAGATAATTTTACTCTAGCCAAAGGATA 656
Qy 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 657 GTGGATGACAGCGGATCCACAGCTTTATGAACATGAACACAGTCCCGCAATGTCGAT 716
Qy 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 717 ATCTATAAAAACTGTACACAGTGTGCTGTCTTCAAAGCAGTGGTTCCTTTACGC 776
Qy 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 777 TGTTTAGCCTGCGGGTCAACTTGAACCTCAAGCGCCGACAGCAGGATCGTGGCGGTGAG 836
Qy 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 837 AGCGCGCTCCCGGGGCGCTGGCCCTGGCAGTGCAGCTGCACGCTCCAGAACGTCACGCTG 896
Qy 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db 897 TGGGAGGCTCCATCATCCCCCGAGTGGATCGTCAGACGCCGCCCACTGCGTGGGAAAA 956
Qy 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 957 CCTCTTAACAATCCATGGCATTTGGCGGCATTTGCGGGGATTTTGAGACATCTTTTCATG 1016
Qy 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 1017 TTCTATGGAGCCGATACCAAGTACAAAAAGTATTTCTCATCCAAATTTATGACTCCAAAG 1076
Qy 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1077 ACCAAGAACCAATGACATTTGGCTGATGAAGCTGCAGAACCTCTGACTTTCAACAGACCTA 1136
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1137 GTGAACACGATGTCTGCCAACCCAGGCATGATGCTGCAGCCAGACAGACTCTGCTGG 1196
Qy 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1197 ATTTCCGGTGGGGGCCACCCGAGGAGAAAGGAGAACCTCAGAGTGTGGAACGCTGCC 1256
Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAsnLeuIle 420
Db 1257 AAGTGTCTCTCATTTGAGACACAGATGCAACAGCAGATATGTCTATGCAACCTGATC 1316
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1317 ACACGACCATGATCTGTGCGGCTTCTGCGAGGAAACGTCGATTTCTTCCAGGGGTGAC 1376
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1377 AGTGGAGGGCTCTGGTCACTTCGAAACAACAATATCTGGTGGCTGTATAGGGGATACAAGC 1436
Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480

Db 1437 TGGGGTCTGGCTGTGCAAGGCTTACAGACCAGGAGTGTACGGGAATGTGATGTGTTATTC 1496
Qy 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1497 ACGACTGGATTATCGACAAATGAGCGCAACCGC 1532

RESULT 5

US-09-685-166A-894
; Sequence 894, Application US/09685166A
; Patent No. 6630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C21
; CURRENT APPLICATION NUMBER: US/09/685.166A
; CURRENT FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 894
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-685-166A-894

Alignment Scores:
Pred. No.: 2,42e-255 Length: 2479
Score: 2696.00 Matches: 486
Percent Similarity: 99.80% Conservative: 5
Best Local Similarity: 98.78% Mismatches: 1
Query Match: 99.23% Indels: 0
DB: 4 Gaps: 0

US-09-615-285B-2 (1-492) x US-09-685-166A-894 (1-2479)
Qy 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 57 ATGGCTTTGAACTCAGGCTACCCACCTATTGGACTTACTATGAAACCATGGATAC 116
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 117 CAACCGGAAACCCCTATCCCGCACAGCCACTGTGGTCCCACTGTCTACGAGGTGCAT 176
Qy 41 ProAlaGlnTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 177 CCGGCTAGTACTACCCGTCCTCCCGTGCAGTACGAGGAGGCTTGCAGCAGGCT 236
Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 237 TCCAAACCCCGTCTGTCAGCAGCCCAATCCCATCCCGGACAGTGTGCACCTCAAG 296
Qy 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 297 ACTAAGAAAGCACTGTGATCACCTTGACCTGGGGACCTTCTTCGTGGGAGCTGCGCTG 356
Qy 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120

Db 357 GCGCGTGGCCCTACTCTGTGAAGTTTCATGGCGAGCAAGTGTCTCCAACCTCTGGATGAGTGC 416
Qy 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
Db 417 GACTCTCTCAGGTACTGTGCATCAACCCCTCTAACTGGTGTGATGGCGTGTACACTGCGCC 476
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 477 GCGCGGAGGAGCAGATCGGTGTGCTTCCGCTCTACGGACCAAACTTCTCTTCAGATG 536
Qy 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Db 537 TACTCATCTCAGAGGAGTCTGGCACCCCTGTGTGCCAAGACGACTGGAACGAGAACTAC 596
Qy 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 597 GCGCGGCGGCTCGCAGGACATGGGCTATAGGAATAATTTTACTCTAGCCALAGGAATA 656
Qy 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 657 GTGGATGACAGCGGATCCACGAGCTTTATGAACACACAGTGCAGGCAATGTGAT 716
Qy 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 717 ATCTATAAAAACTGTACACAGTGCCTGTTTTCANAAAGCAGTGGTTCCTTTACGC 776
Qy 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 777 TGTATTAGCTCGCGGCTCACTTGAACACAGCCGACAGCAGGATCGTGGCGGTGAG 836
Qy 261 SerAlaLeuProGlyAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
Db 837 AGCGCGCTCCCGGGGCGCTGCGCTGGCAGGTGACGCTGCACGCTCCAGAACGTCCACGTG 896
Qy 281 CysGlyGlySerIleIleThrProGlnTrpIleValThrAlaAlaHisCysValGluLys 300
Db 897 TCGCGAGGCTCCATCATCACCCCGGAGTGGATCGTACAGCGCCGCTGCTGGTGAAGAAA 956
Qy 301 ProLeuAsnAsnProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 957 CCTCTTAACATCCATGTCATTTGGACGCGCATTTGGCGGGATTTTGAGACAATCTTTTCATG 1016
Qy 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 1017 TTCTATGAGCGCGATACCAAGTACAAAGATGATTTCTCATCCAAATATGACTCCCAAG 1076
Qy 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1077 ACCAAGAACATGACATTCGCTGATGAAGCTGCAGAAAGCCTCTGACTTTCAACGACCTA 1136
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1137 GTGAACCCAGTGTGTGCCCCAACCCAGGCAATGATGCTGCAGCCAGAACAGCTCTGCTGG 1196
Qy 381 IleSerGlyTyrGlyValAlaThrGluGluLysGlyLysThrSerGlyValLeuAsnAlaAla 400
Db 1197 ATTTCCGGTGGGGGCCACCGAGAGAAAGGAGAGCCTCAGAAAGTGTGAACGCTGCC 1256
Qy 401 LysValLeuLeuIleLeuThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1257 AAGGTGCTTCTCATTGAGACACAGAGATGCAACAGCAGATATGTCTATGACAACTGATC 1316
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1317 ACACAGGCGCATGATCTGTGCGCGCTTCTGCGAGGAGAGCTGATTCCTTCCAGGGTGAC 1376
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1377 AGTGGAGGCGCTCTGGTCACTTCGAACCAACAATATCTGGTGGCTGATAGGGATACAAGC 1436
Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480

Db 1437 TGGGGTTCCTGGCTGCCAAAGCTTACAGACCAGGAGTGTACGGGATGTGTATTC 1496
Qy 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
Db 1497 ACGGACTGGATTTATCGACAAATGAAGCAACGGC 1532

RESULT 6
US-08-807-151-2
; Sequence 2, Application US/08807151
; Patent No. 6043033
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,151
; FILING DATE: Filed Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PP-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1077 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNOT01
; CLONE: 556016
US-08-807-151-2

Alignment Scores:
Pred. No.: 5,788-145 Length: 1077
Score: 1569.00 Matches: 288
Percent Similarity: 99.65% Conservative: 0
Best Local Similarity: 99.65% Mismatches: 1
Query Match: 57.75% Indels: 0
DB: 3 Gaps: 0

US-09-615-285B-2 (1-492) x US-08-807-151-2 (1-1077)

Qy 204 SerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAspIleTyrLys 223
Db 83 ACGGATCCACGAGCTTTATGAAATGAACACAGTGGCGGCAATGTCATATAA 142

Qy 224 LysLeuTyrHisSerAspAlaCysSerLysAlaValSerLysArgCysIleAla 243
Db 143 AAATGTACACAGTGGCTGTTCTTCAAAAGCAGTGGTTCCTTTACGCTGTATAGCC 202

Qy 244 CysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlySerAlaLeu 263

Db 203 TGGGGGTCAACTGAACCTCAAGCCGCCAGAGCAGGATCGTGGCGGCGAGAGCGGCTC 262
Qy 264 ProGlyValaTyrProTyrGlnValSerLeuHisValGlnAsnValHisValCysGlyGly 283
Db 263 CCGGGGGCGCTGGCCCTGGCAGGTGACCTGCAGTCCAGACGTCACCGTGTGGGAGGC 322

Qy 284 SerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGlnLysProLeuAsn 303
Db 323 TCCATCATCACCCCGAGTGGATCGTGACAGCCGCCCACTGCGTGGAAACACCTCTAAC 382

Qy 304 AsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGly 323
Db 383 ATCCATGGCATTTGGACGGCATTTGGGGGATTTGGAGCAATCTTTCATGTTCTATGA 442

Qy 324 AlaGlyTyrGlnValGlnLysValIleSerHisProAsnTyrAspSerLysThrLysAsn 343
Db 443 GCGGATACCAAGTAGAAAAGTGAATTTCTCATCCAAATATGACTCCAGACCAAGAAC 502

Qy 344 AsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuValLysPro 363
Db 503 AATGACATTCGCGCTGATGAGCTGCAAGAGCTTCTGACTTTTCAACGACCTAGTGAACCA 562

Qy 364 ValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyrIleSerGly 383
Db 563 GTGTGCTGCCCAACCCAGGATGATGCTGCAGCCAGAACAGCTCTGCTGGATTTCCGGG 622

Qy 384 TrrGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaLysValLeu 403
Db 623 TGGGGGGCCACCGAGGAGAAAGGAGACCTCAGAAGTGTCTGAACCTGCAAGGTGCTT 682

Qy 404 LeuIleLthrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIleThrProAla 423
Db 683 CTCAATTGAGACACAGAGTGCACACAGACAGATATGCTATGACAACTGATCACACGCC 742

Qy 424 MetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyGly 443
Db 743 ATGATCTGTGCGGCTTCCTTGCAGGGGAAACGTCGATTTCTTGCAGGGTGACAGTGGAGG 802

Qy 444 ProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSerTrpGlySer 463
Db 803 CNTCTGTCCTTCGAAAGAACATATCTGGTGGCTGATAGGGGATACAGCTGGGTCT 862

Qy 464 GlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPheThrAspTyr 483
Db 863 GGCTGTGCCAAAGCTTACAGACCAGGAGTGTACGGGATGTGTGATGTATTCACGGACTGG 922

Qy 484 IleTyrArgGlnMetArgAlaAspGly 492
Db 923 ATTTATCGACAAATGAGGGCAGACGGC 949

RESULT 7
US-09-478-957-2
; Sequence 2, Application US/09478957
; Patent No. 6350448
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/09/478,957
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/807,151
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1077 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORN0T01
; CLONE: 556016
; US-09-478-957-2

Alignment Scores:
Pred. No.: 5,78e-145 Length: 1077
Score: 1569.00 Matches: 288
Percent Similarity: 99.65% Conservativeness: 0
Best Local Similarity: 99.65% Mismatches: 1
Query Match: 57.75% Indels: 0
DB: 4 Gaps: 0

US-09-615-285b-2 (1-492) x US-09-478-957-2 (1-1077)
QY 204 SerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAspIleTyrLys 223
DB 83 AGCGATCCACAGCTTTATGAACCTGACACAGGTCGGGCAATGCTATCTATATAA 142
QY 224 LysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArgCysIleAla 243
DB 143 AAACGTGTACACAGTGTGCTCTTCTCAAAAGCAGTGTGTTCTTACGCTGTATAGC 202
QY 244 CysGlyValAsnLeuAsnSerArgGlnSerArgIleValGlyGlyGluSerAlaLeu 263
DB 203 TCGGGGTCAACTTGAACTCAACCCGACGAGCAGGATGTGGGGCGGCGAGCGGCTC 262
QY 264 ProGlyValTrpProTrpGlnValSerLeuHisValGlnAsnValHisValCysGlyGly 283
DB 263 CCGGGGCGCTGGCCCTGGCAGGTGAGCTGCAGTCCAGAACGTCACGCTGTGGGAGGC 322
QY 284 SerIleIleThrProGluTolIleValThrAlaAlaHisCysValGluLysProLeuAsn 303
DB 323 TCCATCATCACCCCGAGTGGATCGTGACAGCCGCCCACTGCGTGGAAAACCTCTTAAC 382
QY 304 AsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGly 323
DB 383 AATCCATGGCATTTGGACGGCATTTGGGGGATTTTGAGACAATCTTTCATGTTCTATGA 442
QY 324 AlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLysThrLysAsn 343
DB 443 GCGGGATACCAAGTAGAAAAGTGAATTTCTATCCAAATTTATGACTCCAAAGCCAGAAC 502
QY 344 AsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuValLysPro 363
DB 503 AATGACATTCGGCTGTATGAAGCTGACAGACCTCTGACTTTCACACACCTAGTGAACCA 562
QY 364 ValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrpIleSerGly 383
DB 563 GTGTGTCTGCCCAACCCAGGCGATGATGCTGACGCCAGAACAGCTCTGTCTGGATTTCCGGG 622
QY 384 TrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaLysValLeu 403
DB 623 TGGGGGGCCCAACCGAGGAGAAAGGAGACCTCAGAAGTGTCTGAACGCTGCCAAGTGCTT 682
```

```

QY 404 LeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIleThrProAla 423
DB 683 CTCATTGAGACACAGATGCAACAGCAGATATGCTATGACACACCTGATCACACAGCC 742
QY 424 MetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyGly 443
DB 743 ATGATCTGTGCGGCTTCTTCTGACGGGAAACGTCGATTTCTGCCAGGGTGACAGTGGAGG 802
QY 444 ProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSerTrpGlySer 463
DB 803 CNTCTGTGCTACTCGAAGAACATATCTGTGGTGTGATAGGGGATACAGCTGGGGTCT 862
QY 464 GlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPheThrAspTrp 483
DB 863 GGCTGTGCCAAAGCTTACACACAGGAGTGTACGGGAATGTGATGTATTACGAGACTGG 922
QY 484 IleTyrArgGlnMetArgAlaAspGly 492
DB 923 ATTTATCGACAAATGAGGCGACGCGC 949

RESULT 8
US-09-016-434-963
; Sequence 963, Application US/09016434
; Patent No. 650938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 963:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT05
; CLONE: 842889
; US-09-016-434-963

Alignment Scores:
Pred. No.: 1.2e-136 Length: 1001
Score: 1484.00 Matches: 279
Percent Similarity: 94.61% Conservativeness: 2
```


Best Local Similarity: 93.94% Mismatches: 12
Query Match: 54.62% Indels: 4
DB: 4 Gaps: 0

US-09-615-285B-2 (1-492) x US-09-016-434-963 (1-1001)

```
QY 1 MetAlaLeuAsnSerGlySerProProAlaileGlyProTyrTyrGluAsnHisGlyTyr 20
Db 109 ATGGCTTTGAACCTCAGGCTCACCACCGCTATTGGACCTTACTATGAAACCACTGGATAC 168
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 169 CAACCGGAAACCCCTATCCCGCACGCCACTGTGTGCCACTGTCTACGAGGTGCAT 228
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 229 CCGGCTGACTACTACCGCTCCCGCTGCCAGTACGCCCGAGGCTCTTACGCGAGCT 288
QY 61 SerAsnProValValCysThrGlnProLysSerPro--SerGlyThrValCysThrSerL 80
Db 289 TCCAACCCGCTGCTGCACGACGCCCAATCCCATCTCKGACNTGTGTGCACCTCAA 348
QY 80 YsThrLysLeuAlaLeuCysLeuThrLeuThrLeuGlyThrPheLeuValGlyValAlaL 100
Db 349 AGACTAAGAAAGCACTGTGCATCACCTTGACCTTGGGACCTTCTCTGTGGAGCTGCC 408
QY 100 euAlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyLeuGluC 120
Db 409 TGGCGCTGGCTACTCTGGAAGTTTCATGGCAGCAAGTGTCTCAACTCTGGGATAGT 468
QY 120 YsAspSerSerGlyThrCysLeuAsnProSerAsnTyrCysAspGlyValSerHisCysP 140
Db 469 GCGACTCTCAGGTACCTGCATCAACCCCTCTAAGTGTGTGGTGTGTGCACCTGCC 528
QY 140 roGlyGlyGlu-AspGluAsnArgCysValArgLeuTyrGlyProAsnPheLeuLeuGln 159
Db 529 CCGCGGGGAGGACAAATCGGTGTTTCGCTCTACGACCAAACTTCATCTTCAG 588
QY 160 ValTyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTyrAsnGluAsn 179
Db 589 GTGTACTCATCTCAGAGAACTCTCGCACCTGTGTGCCAAGACNACTGAAACGAGAAC 648
QY 180 TyrGlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGly 199
Db 649 TACGGCGGGGGCTCGCAGGACATGGCTATTAAGATTAATTTTACTCTAGCCAAAGGA 708
QY 200 IleValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnVal 219
Db 709 ATAGTGATGACAGCGGATCNCAGGCTTTATGAACCTGAACACAGTGCAGGCAATGTC 768
QY 220 AspileTyrLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeu 239
Db 769 GATATCTATAAAACATGTACACAGTATGCCCTGTTCTTCAAAGCAGTGGTTCTTTA 828
QY 240 ArgCysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGly 259
Db 829 CGCTGTATNCTTGGGGGTCACTTGAACCTAAGCGCCAGAACAGAGATCGTGGGGGCG 888
QY 260 Glu-SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHi 279
Db 889 GAANAGCGCGCTCCGGGGGCTGGCCCTGGCAGGTGAGCCCTGCAGCTCCAGAACTCCAC 948
QY 279 svalCysGlyGlySerIleThrProGluTyrIleValThr 293
Db 949 NTTTTGCGAAGCTCCATCATCAACCCCGAATTGATCGTGACA 991
```

RESULT 9

US-09-685-166A-896
; Sequence 896, Application US/09685166A
; Patent No. 6630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangshun
; APPLICANT: Dillon, Davin C.

APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
TITLE OF INVENTION: DIAGNOSIS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427C21
CURRENT APPLICATION NUMBER: US/09/685,166A
NUMBER OF SEQ ID NOS: 898
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 896
LENGTH: 683
TYPE: DNA
ORGANISM: Homo sapiens
US-09-685-166A-896

Alignment Scores:
Pred. No.: 1,67e-105 Length: 683
Score: 1165.00 Matches: 208
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.52% Mismatches: 0
Query Match: 42.88% Indels: 0
DB: 4 Gaps: 0

US-09-615-285B-2 (1-492) x US-09-685-166A-896 (1-683)

```
QY 1 MetAlaLeuAsnSerGlySerProProAlaileGlyProTyrTyrGluAsnHisGlyTyr 20
Db 57 ATGGCTTTGAACCTCAGGCTCACCACGACTATGGACCTTACTATGAAACCACTGGATAC 116
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 117 CAACCGGAAACCCCTATCCCGCACGCCACTGTGTGCCACTGTCTACGAGGTGCAT 176
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 177 CCGGCTCAGTACTACCGCTCCCGCTGCCAGTACGCCCGGAGGCTCTGACGAGGCT 236
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 237 TCCAACCCCGCTGTCTGCACGCGACGCCCAATCCCATCCGCGACAGTGTGCACCTCAAAG 296
QY 81 ThrLysLeuAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
Db 297 ACTAAGAAAGCACTGTGCATCACCTTGACCTTGGGAGACCTTCTCTGTGGAGCTGCGTGC 356
QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 357 GCGGCTGGCTACTCTGGAGTTCATGGCACCAAGTGTCTCCACTCTGGGATAGAGTGC 416
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
Db 417 GACTCTCAGGTACCTGCATCAACCCCTCTAAGTGTGTGTGGGTGTGCACACTGCCCTC 476
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 477 GCGGGGAGGACGAGAAATCGGTGTGTTGCCCTCTACGGACCAAACTTCATCTCTCAGATG 536
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
Db 537 TACTCATCTCAGAGGAAGTCTCTGGCACCCCTGTGTGCCAAGACGACTGGAAACGAGACTAC 596
```

QY 181 GlyArgAlaAlaCysArgSerMetGlyTyrLysAsnAsnPhetTyrSerSerGlnGlyLe 200
Db 597 GGGCGGGCGGCTGACAGGACATGGGCTATAGAAATAATTTTACTCTAGCCAAAGGAATA 656
QY 201 ValAspAspSerGlySerThrSerPhe 209
Db 657 GTGGATGACAGCGGATCCACCGACTTT 683

RESULT 10

US-09-518-046-1
; Sequence 1, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CJP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 1
; LENGTH: 2413
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; OTHER INFORMATION: entire cDNA sequence of TADG-12 gene
US-09-518-046-1

Alignment Scores:

Pred. No.: 2,248-77 Length: 2413
Score: 888.00 Matches: 205
Percent Similarity: 53.06% Conservative: 64
Best Local Similarity: 40.43% Mismatches: 186
Query Match: 32.68% Indels: 52
DB: 3 Gaps: 15

US-09-615-285B-2 (1-492) x US-09-518-046-1 (1-2413)

QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrVal----- 36
Db 24 AAGCCAGTAACACTGTGGCTACTATCTCTCCGTGGTGCATCTACATTTTGGGACTC 83
QY 37 -----TyrGluValHisProAlaGlnTyrTyrProSerProValProGlnTyrAlaPro 54
Db 84 GGGAAATTATGAGTAGAG-----GTGGAGGCGGAGCCGGATGTCAGAGTCTCT 131
QY 55 ArgValLeuThrGlnAlaSerAsn-----ProValVal----- 65
Db 132 GAAATAGTCACCATGGGGGAAATGATCCGCTCTGTGTGAAGCCCTTCTCATTCGGA 191
QY 66 -----CysThrGlnProLysSerProSerGlyThrValCys 77
Db 192 TCGCTTTTGGCTTGATGATTGAAATAAGTCTGTGTGCACCATGACAGATGCTGTT 251
QY 78 ThrSerLysThrLysLysAlaLeuCysIleThrLeu-----ThrLeuGly 92
Db 252 GGTGCACAGATCT 311
QY 93 ThrPheLeuValGlyAlaAlaLeuAlaGlyLeuLeuTyrLysPheMetGlySerLys 112
Db 312 ATCATTTGATGATATTAGCACTGGCCATGCTGTGGGATCCACTTC-----GAC 362
QY 113 CysSerAsnSerGlyIleGluCysAspSerSerGlyThrCysIleAsnProSerAsnTrp 132
Db 363 TGCTCAGGGAAG---TACAGATGTGCTCATCTCTTAAGTGTATCGAGCTGATAACTCGA 419
QY 133 CysAspGlyValSerHisCysProGlyGlyGluAspGluAsnArgCysValArgLeuTyr 152

Db 420 TGTGACGAGTCTCGGATTGCAAGACGGGAGACGAGTACCGCTGTGTCCGGTGGGT 479
QY 153 GlyProAsnPhelLeuGlnValTyrSerSerGlnArgLysSerTrpHisProValCys 172
Db 480 GGTCTAGATCGGCTGCTCAGGTGTTCACAGCT-----GCTTCGTGAAAGACCATGTGC 533
QY 173 GlnAspAspTyrAsnGluAsnTyrGlyArgAlaAlaCysArgAspMetGlyTyrLysAsn 192
Db 534 TCCGATGACTGGAAGGTCCTAGCAAAATGTTGCTGTGCCCACTGGGTTCCTCA--- 590
QY 193 AsnPhetTyrSerSerGlnGlyIleValAspAspSerGlySerThrSerPheMetLysLeu 212
Db 591 AGCTATGTGAGTTTTCAGATAACTCAGAGTGCCTGCTGGAGGGGACAGTTCCGGAGAGAG 650
QY 213 AsnThrSerAlaGlyAsnVal-----AspIleTyrLysLysLeuTyrHisSer--- 228
Db 651 TTTGTGTCTCATCGATCACCTCTTCCAGATGACAGAGTGCATGCAATTACACCACTCAGTA 710
QY 229 -----AspAlaCysSerSerLysAlaValSerLeuArgCysIleAlaCysGly 245
Db 711 TATGTGAGGGAGGATGTGCTCTGGCCACGTGTACCTTGCAGTGCACAGGCTGTGGT 770
QY 246 ValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGluSerAlaLeuProGly 265
Db 771 CATGAAGGGGCTACAGC-----TCACGCATCGTGGTGGAAACATGCTCTTCTCTCG 824
QY 266 AlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisValCysGlyGlySerIle 285
Db 825 CAGTGGCCCTGGCAGGCGGACGCTTCAGTTCAGGGCTTACCACCTGTGGGGCTCTGTCTC 884
QY 286 IleThrProGluTyrIleValThrAlaAlaHisCysValGluLysProLeuAsnAsnPro 305
Db 885 ATCAGCCCTCTGTGATCATCATCTGCTGCACACTGTGTT---TATGACTGTGACCTCCCC 941
QY 306 TrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGlyAla--- 324
Db 942 AAGTCATGACCATTCAGGTGGGTCTAGTT-----TCCCTGTGGCAATTCACAGCCCCA 995
QY 325 GlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLysThrLysAsnAsn 344
Db 996 TCCCACTGTGTGGAGAGATGTGTACACAGCAAGTACAAAGCAAGAGGCTGGGCAAT 1055
QY 345 AspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuValLysProVal 364
Db 1056 GACATCGCCCTTATGAAGCTGGCGGGCCACTCAGCTTCAATGAAATGATCCAGCCTGTG 1115
QY 365 CysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrpIleSerGlyTyr 384
Db 1116 TGCCTGCCACTCTGAAGAGAACTTCCCGCATGGAAAGTGTGTGGACGTGAGGATGG 1175
QY 385 GlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAlaLysValLeuLeu 404
Db 1176 GGGGCCACAGAGGATGAGGTGAGCGCTCCCTGTCTGAACCAACCGCGCCGCTCCCTTTG 1235
QY 405 IleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIleThrProAlaMet 424
Db 1236 ATTTCCAAAGATCTGCAACCAACAGGAGCTGTACGGTGGCATCATCTCCCTCCCATG 1295
QY 425 IleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnClyAspSerGlyLysPro 444
Db 1296 CTCGCGCGGCTACCTGACGGGTGGCGTGAACAGCTGCCAGGGGACAGCGGGGGCCCC 1355
QY 445 LeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSerTrpGlySerGly 464
Db 1356 CTGGTGTCTCAAGAGAGGAGGCTGTGGAGGTAGTGGAGCGGACAGCTTTGGCATCGGC 1415
QY 465 CysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPheThrAspTrpIle 484
Db 1416 TGCCAGAGGTGAACAGCTGGGGGTACACCCCGTGTCACTCTCTCTGACTGGATC 1475
QY 485 TyrArgGlnMetArgAlaAsp 491
Db 1476 CACGAGCAGATGGAGAGAGAC 1496

RESULT 11

US-09-518-046-3
 ; Sequence 3, Application US/09518046
 ; Patent No. 6294663
 ; GENERAL INFORMATION:
 ; APPLICANT: O'Brien, Timothy J.
 ; APPLICANT: Underwood, Lowell J.
 ; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
 ; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
 ; FILE REFERENCE: D6192CIP
 ; CURRENT APPLICATION NUMBER: US/09/518,046
 ; CURRENT FILING DATE: 2000-03-02
 ; EARLIER APPLICATION NUMBER: 09/261,416
 ; EARLIER FILING DATE: 1999-03-03
 ; NUMBER OF SEQ ID NOS: 153
 ; SEQ ID NO 3
 ; LENGTH: 2544
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; OTHER INFORMATION: entire cDNA sequence of TADG-12 variant gene
 US-09-518-046-3

Alignment Scores:

Pred. No.: 6,29e-75 Length: 2544
 Score: 863.50 Matches: 208
 Percent Similarity: 49.27% Conservative: 63
 Best Local Similarity: 37.82% Mismatches: 185
 Query Match: 31.78% Indels: 95
 DB: 3 Gaps: 16

US-09-615-285B-2 (1-492) x US-09-518-046-3 (1-2544)

```

QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrVal----- 36
DB 24 AAGCCAGTAACACTGTGGCCCTACTATCTTCCGCTGGTGCCATCTACATTTTGGGACTC 83
QY 37 -----TyrGluValHisProAlaGlnTyrTyrProSerProValProGlnTyrAlaPro 54
DB 84 GGGAAATTATGAGTAGAG-----GTGGAGCGCGGCGGATGTGCAGAGGTCCT 131
QY 55 ArgValLeuThrGlnAlaSerAsn-----ProValVal----- 65
DB 132 GAATAGTACACATGGGGGAAATGATCCGCTCTCTGTTGAGCCCTTCTCATTCGGA 191
QY 66 -----CysThrGlnProLysSerProSerGlyThrValCys 77
DB 192 TCGCTTTTGGCCTTGATGTTGAAATAAGTCTCTGTGACCAGATGCAGATGCTGTT 251
QY 78 ThrSerLysThrLysLysAlaLeuCysIleThrLeu-----ThrLeuGly 92
DB 252 GCTGCACAGATCCTGTCACTGCTGCCATTTGAAAGTTTTTCCTCCCAATCATGCTCATTTGGG 311
QY 93 ThrPheLeuValGlyAlaAlaLeuAlaAlaGlyLeuLeuTyrLysPheMetGlySerLys 112
DB 312 ATCATGTCATTGATATTAGACTGGCCATCTGTTGGGCATCCACTTC-----GAC 362
QY 113 CysSerAsnSerGlyIleGluCysAspSerSerGlyThrCysIleAsnProSerAsnTrp 132
DB 363 TGCTCAGGGAAG---TACAGATGTGCTCATCTCTTTTAAGTGTATCGAGCTGATTAATCGA 419
QY 133 CysAspGlyValSerHisCysProGlyGlyGluAspGluAsnArgCysValArgLeuTyr 152
DB 420 TGTGACGAGTCTCGGATTGCAAGACGGGAGGAGGATACCCGCTGTGTCCGGTGGGT 479
QY 153 GlyProAsnPheIleLeuGlnValTyrSerSerGlnArgLysSerTrpHisProValCys 172
DB 480 GGTCAAGATCGCGTCTCCAGGTGTTCACAGCT-----GCTTCGTGGAAGACCATGTGC 533
QY 173 GlnAspAspTrpAsnGluAsnTyrGlyArgAlaAlaCysArgAspMetGlyTyrLysAsn 192

```

```

DB 534 TCCGATGACTGGAGGGTCACTACGCAATGTTGCTGTGCCCAACTGGTGGTTTCCCA--- 590
QY 193 AsnPheTyrSerSerGlnGlyIleValAspAspSerGlySerThrSerPheMetLysLeu 212
DB 591 AGCTATGTAAGTTTCAGATAAAGTCAAGAGTCTGCTGGAGGGGAGTTCGGGGAGGAG 650
QY 213 AsnThrSerAlaGlyAsnVal-----AspIleTyrLysLysLeuTyrHisSer--- 228
DB 651 TTTGTGTCCATCGATCACCCTTTCAGATGACAAGAGTGAAGTCACTTACACACATCAGTA 710
QY 229 -----AspAlaCysSerSerLysAlaValSerLeuArgCysIleAlaCysGly 245
DB 711 TATGTGAGGGAGGAGTGTCCCTCTGGCCACGTGGTTCCTTCAGTGCACACCTTGTGT 770
QY 246 ValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlySerAlaLeuProGly 265
DB 771 CATAGAAGGGGTACAGC-----TCACGATCGTGGTGGAAACATGTCCTTGTCTCG 824
QY 266 AlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisValCysGlyGlySerIle 285
DB 825 CAGTGGCCCTGGAGCCAGCCTTCAGTTCACAGGCGTACCACCTGTGCGGGGCTCTGTC 884
QY 286 IleThrProGluTyrIleValThrAlaAlaHisCysVal----- 298
DB 885 ATCAGCGCCCTGTGGATCATCACTGTCGACACTGTGTTATGAGATTGTAGTCTCTAGA 944
QY 299 -----GluLysProLeuAsnAsn--- 304
DB 945 GAAAGGGCAGACAGAGAGAGGAAGTCTCTGTCTGGAGGAAACCCAC-AAAAATGAA 1003
QY 304 ----- 304
DB 1004 AGGACCTAGACCTTCCCATAGCTAATTCAGTGGACCATGTTATGGCAGATACAGGCTTG 1063
QY 305 -----ProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyr 322
DB 1064 TACCTCCCAAGTCAATGAGCCATCCAGGTGGGTCTAGTT-----TCCCTGTGGACAA 1117
QY 323 GlyAla---GlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLysThr 341
DB 1118 CAGCCCCATCCCATCTGTGTGGAGAGATTGCTACCAACAGCAGTACAGCCAAAGAGG 1177
QY 342 LysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuVal 361
DB 1178 CTGGGCAATGACATCGCCCTTATGAAGTGGCGGGCCACTCACGTTCAATGAATGATC 1237
QY 362 LysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrpIle 381
DB 1238 CAGCCTGTGTGCTGCCCAACTCTGAAGAGAACTTCCCGCATGGAAGAGTGTGGAGC 1297
QY 382 SerGlyTyrGlyValaThrGluGlyLysThrSerGluValLeuAsnAlaAlaLys 401
DB 1298 TCAGATGGGGGGCCACAGAGATGAGGTGACGCTCCCTGTCTGTAACCCACCGGCGC 1357
QY 402 ValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIleThr 421
DB 1358 GTCCCTTTGATTTCAACAAGATCTGCAACACAGGGAGCTGTACGGTGGCATCATCTCC 1417
QY 422 ProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAspSer 441
DB 1418 CCTCCATGCTCTGGCGGGCTACTGTGCGGTGGCGTGGACAGCTGCCAGGGGACAGC 1477
QY 442 GlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSerTrp 461
DB 1478 GGGGGGCCCTGTGTGTCAAGAGAGGAGGCTGTGGAAAGTTAGTGGGAGCGACAGCTTT 1537
QY 462 GlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPheThr 481
DB 1538 GGATCGGTGGCAGAGGTGAACAGCTGGGGGTGTACCCCGTGTACCTCCCTCTCTG 1597
QY 482 AspTrpIleTyrArgGlnMetArgAlaAsp 491
DB 1598 GACTGGATCCACGAGCAGATGGAGAGAC 1627

```

```
RESULT 12
US-09-261-416-1
; Sequence 1, Application US/09261416A
; Patent No. 6291563
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TAGD-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261.416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: 144..1511
; OTHER INFORMATION: CDS
US-09-261-416-1

Alignment Scores:
Pred. No.: 2,02e-74 Length: 2416
Score: 858.00 Matches: 206
Percent Similarity: 52.75% Conservative: 63
Best Local Similarity: 40.39% Mismatches: 186
Query Match: 31.58% Indels: 55
DB: 3 Gaps: 15

US-09-615-285B-2 (1-492) x US-09-261-416-1 (1-2416)
Qy 21 GlnProGluAenProTyProAlaGlnProThrValProThrVal----- 36
Db 24 AAGCCAGTAACACTGCGCCCTACTATCTCTCCGTGGTCCCATCATATTTGGGACTC 83
Qy 37 -----TyrGluValHisProAlaGlnTyTyProSerProValProGlnTyAlaPro 54
Db 84 GGGNAATATGAGGTAGAG-----GTGGAGGCGGAGCGGATGTCAGAGTCT 131
Qy 55 ArgValLeuThrGlnAlaSerAen-----ProValVal----- 65
Db 132 GAAATAGTACCACCTGGGGGAAATGATCCGCTCTGTTGAAGCCCTTCTCATTCGGA 191
Qy 66 -----CysThrGlnProLysSerProSerGlyThrValCys 77
Db 192 TCGCTTTTGGCCTTGATGATTGAAATAGTCTCTGTTGCCACAGATGCGATGCT 251
Qy 78 ThrSerLysThrLysLysAlaLeuCysAlaLeu-----ThrLeuGly 92
Db 252 GCTGCACAGATCCTGCTCACTGCTGCCATTTGAAGTTTTCCTCCCAATCATCGTCA 311
Qy 93 ThrPheLeuValGlyAlaAlaLeuAlaGlyLeuLeuTrpLysPheMetGlySerLys 112
Db 312 ATCATTCGATGATATTAGCACGCGCCCATTTGGTCTGGGCAATCCACTTC-----GAC 362
Qy 113 CysSerAenSerGlyLleGluCysAaspSerSerGlyThrCysLleAenProSerAenTrp 132
Db 363 TGCTCAGGAG--TACAGATGTCGCTCATCTTTAAGTGATTCGAGCTGATAACTCGA 419
Qy 133 CysaspGlyValSerHisCysProGlyGlyGluaspGluaspGluaspGluaspGluasp 152
Db 420 TGTGACGGAGTCTCGGATTCGAAAGACGGGGAGGACGAGTACCGCTGTGTCGGGGTGG 479
Qy 153 GlyProAenPheLeuLeuGlnValTySerSerGlnArgLysSerTrpHisProValCys 172
Db 480 GGTGAGAAATGCGTCTCCAGGTGTTCCACAGCT-----CCTTCGTGGAGACCATGTGC 533
Qy 173 GlnAaspTrpAenGluAenTyTyGlyArgAlaAlaCysArgAaspMetGlyTyTyLysAen 192
Db 534 TCCGATGACTGGAAGGCTCACTACGCAATGTTGCTGTGCCAACTGGGTTTCCCA---- 590
```

```
Qy 193 AsnPheTySerSerGlnGlyLleValAaspAaspSerGlySerThrSerPheMetLysLeu 212
Db 591 AGCTATGTGAGTTCAGATAACCTCAGAGTACGCTCGTGGAGGGGACGATTCGGGGAGGAG 650
Qy 213 AsnThrSerAlaGlyAenVal-----AspileTyLysLysLeuTyHiser--- 228
Db 651 TTTGTGTCATCATCATCCTCTTCCAGATGACAGGTGATGATACACCATCTCAGTA 710
Qy 229 -----AspAlaCysSerSerLysAlaValValSerLeuArgCysLleAlaCysGly 245
Db 711 TATGTGAGGAGGAGGATGTCCTCTGGCCACGCTGTACTCTTCAGTGCAGAGCTGTGTGT 770
Qy 246 ValAenLeuAenSerSerArgGlnSerArgLleValGlyGlyGlyGluSerAlaLeuProGly 265
Db 771 CATGAGAGGGGCTACAGC-----TCACGATCTGGTGGTGAACATGCTCTCTGCTCTCG 824
Qy 266 AlaTrpProTrpGlnValSerLeuHisValGlnAenValHisValCysGlyGlySerile 285
Db 825 CAGTGGCCCTGGCAGGCGCAGCTTCAGTTCAGGGCTACACCTGTGGGGGCTCTGTC 884
Qy 286 IleThrProGluTrpLleValThrAlaAlaHisCysValGluLysProLeuAenAenPro 305
Db 885 ATCAGCCCTCTGTGATCATCTGCTGCACACTGTGT---TATGACTTGTACCTCCTCCC 941
Qy 306 TrpHisTrpThrAlaPheAlaGlyLleLeuArgGlnSerPheMetPheTyTyGlyAla--- 324
Db 942 AAGTCATGACCATCCAGGTGGTCTAGT-----TCCCTGTGGACATCCAGCCCA 995
Qy 325 GlyTyGlnValGluLysValLleSerHisProAenTyTyAspSerLysThrLysAenAen 344
Db 996 TCCACATTTGGTGGAGAGATTGTTTACCCACAGCAAGTACAAAGCCCAAGAGCTGGGCAAT 1055
Qy 345 AspLleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAaspLeuValLysProVal 364
Db 1056 GACATCGCCCTTATGAGTGGCGGGCCACTCAGCTTCATGAAATGATCCAGCTGTG 1115
Qy 365 CysLeuProAenProGlyMetMetLeuGlnProGluGlnLeuCysTrpLleSerGlyTrp 384
Db 1116 TGCCTGCCCACTCTGAAGAGAACTTCCCGCATGGAAAAGTGTCTGGACGTCAAGATGG 1175
Qy 385 GlyAlaThrGluGluLysGlyLysThrSerGluValLeuAenAlaAlaLysValLeuLeu 404
Db 1176 GGGGCCACAGAGATGGAGGTGAGCTCCCTCTGTAACACCGCGCGCTCCCTTTG 1235
Qy 405 IleGlu-ThrGlnArgCysAenSerArgTyTyValTyAaspAenLeuLeuThrProAlaMe 424
Db 1236 ATTTCCAAAGATCTGCAACACAGGACGTGTACGGTGGCATCATCTCCCTCCCAT 1295
Qy 424 TlleCysAlaGlyPheLeuGlnGlyAenVal--AaspSerCysGlnGlyAaspSerGlyGly 443
Db 1296 GCTTCGCGCGGCTACCTGACGGGTGGGCTGGAAACAGCTCCAGGGGAGCAGCGGGGG 1355
Qy 444 ProLeuValThrSerLysAenAenLleTrpTrpLleLleGlyAaspThrSerTrpGlySer 463
Db 1356 CCCCTGTGTGTCAAGAGAGGAGGCTGTGGAAGTTAGTGGAGCGACCCAGCTTTGGCATC 1415
Qy 464 GlyCysAlaLysAlaTyArgProGlyValTyGlyAenValMetValPheThrAaspTrp 483
Db 1416 GGTTCGGCAGACGTGAACACAGCTGGGTGTACACCGGTGTACACCTCTCTCTGACTGG 1475
Qy 484 IleTyArgGlnMetArgAlaAasp 491
Db 1476 ATCCACGACGATGGAGAGAGAC 1499
```

RESULT 13

```
US-09-656-002-1
; Sequence 1, Application US/09656002
; Patent No. 6455668
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
```

FILE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS

FILE REFERENCE: A-69108/DJB/JDD/AMS

CURRENT APPLICATION NUMBER: US/09/656,002

PRIOR FILING DATE: 2000-03-06

PRIOR APPLICATION NUMBER: US 09/525,993

PRIOR FILING DATE: 2000-03-15

PRIOR APPLICATION NUMBER: US 09/493,444

PRIOR FILING DATE: 2000-01-28

PRIOR APPLICATION NUMBER: PCT/US 00/07044

PRIOR FILING DATE: 2000-03-15

NUMBER OF SEQ ID NOS: 3

SOFTWARE: PatentIn version 3.0

SEQ ID NO 1

LENGTH: 2079

TYPE: DNA

ORGANISM: Homo sapiens

US-09-656-002-1

Alignment Scores:

Pred. No.: 9,71e-58 Length: 2079

Scores: 687.50 Matches: 171

Percent Similarity: 48.86% Conservative: 65

Best Local Similarity: 35.40% Mismatches: 170

Query Match: 25.30% Indels: 77

DB: 4 Gaps: 18

US-09-615-285B-2 (1-492) x US-09-656-002-1 (1-2079)

Qy 41 ProAlaGlnTyrProSerProValProGlnTyrAlaProArgValLeuThr----- 58

Db 203 CCAGCATGGTAC-----AGGATCCTCACAGTGATC 232

Qy 59 -----GlnAlaSerAsnProValValCysThrGlnProLysSerProSer 73

Db 233 AACCTCTGAACAGCCTCGATGTAACCC-----TGGCAAAACCCCGTATCCCCA--- 283

Qy 74 GlyThrValCysThrSerLysThrLysLysAlaLeuCysileThrLeuThrLeuGlyThr 93

Db 284 TGGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGCCTGGCGAGT 343

Qy 94 PheLeuValGlyAlaAlaLeuAlaGlyLeuLeuThrLysPheMetGlySerLysCys 113

Db 344 ATCATCTGGTGGTGGTCTCATCAAGGTGATCTGGATTAATATCTAC----- 391

Qy 114 SerAsnSerGlyIleGluCysAspSerSerGlyThrCysIleAsnProSerAsnTrpCys 133

Db 392 -----TTCCCTCTGGGGCAGCCTCTCCACITCATCCCGAGGAGCAGCTGTGT 439

Qy 134 AspGlyValSerHisCysProGlyGlyGluAspGluAsnArgCysValArgLeuThr--- 152

Db 440 GACGAGAGCTGGACTGTCCCTTGGGGAGGACGAGGACACTGTGTCAAGAGCTTCCCC 499

Qy 153 ---GlyPro-----AsnPheIleLeuGlnValTyrSer 162

Db 500 GAAGGGCTCGAGTGGCAGTCCGCTCTCCAGGACCGATCCACACTGCAGGTCTGGAC 559

Qy 163 SerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyrGlyArg 182

Db 560 TCGGCCACAGGAACCTGTCTCTGCTGTTCGACAACTTCACAGAACTCTCGCTGAG 619

Qy 183 AlaAlaCysArgAspMetGlyTyr-----LysAsnAsnPheTyrSerSerGln----- 198

Db 620 ACAGCCTGTAGCAGATGGCTTACAGAGCAAAACCCACTTTCAGAGCTGTGGATGGC 679

Qy 199 -----GlyIleValAspAspSerGlySerThrSerPheMetLysLeuAsn 213

Db 680 CCAGACAGGATCTGGATGTTGTTGAATATCACAGAAACACAGCCAGGAGCTCGCATCGG 739

Qy 214 ThrSerAlaGlyAsnValAspIleTyrLysLysLeuTyrHisSerAspAlaCysSerSer 233

Db 740 AACTCAAGTGG-----CCCTGTCTCTCA 763

Qy 234 LysAlaValSerLeuArgCysileAlaCysGlyValAsnLeuAsnSerArgGln 253

Db 764 GGCTCCCTCGTCTCCCTGCACCTGTCTGCTGTGGAGAGCCTG-----AAGACC 814

Qy 254 SerArgIleValGlyGlyGluSerAlaLeuProGlyAlaTrpProTrpGlnValSerLeu 273

Db 815 CCCGCTGTGTGGTGGGAGGAGGCTCTGTGGATCTTGTGCTTGGCAGTCTGAGATC 874

Qy 274 HisValGlnAsnValHisValCysGlyGlySerIleThrProGluTrpIleValThr 293

Db 875 CAGTACGACAAACAGCAGCTCTGTGAGGAGGAGCATCTGTGAGCCCTGAGTCTCTCAG 934

Qy 294 AlaAlaHisCysValGluLysProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGly 313

Db 935 GCAGCCCACTGCTTACGAAACAT---ACCGATGTGTCACTGGAGGTGGGGCAGC 991

Qy 314 IleLeuArg---GlnSerPheMetPheTyrGlyAlaGlyTyrGlnValGluLysValIle 332

Db 992 TCAGACAAACTGGCAGCTTC-----CCATCCCTGGCTGTGCTGCCAAGATCATC 1039

Qy 333 -----SerHisProAsnTyrAspSerLysThrLysAsnAsnAspIleAlaLeuMet 349

Db 1040 ATCATTGAATTCACCCCATGTAC-----CCCAAGACAAATGACATGCCCTCATG 1090

Qy 350 LysLeuGlnLysProLeuThrPheAsnAspLeuValLysProValCysLeuProAsnPro 369

Db 1091 AAGCTGCACTTCCCACTCACTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTCTTT 1150

Qy 370 GlyMetLeuGlnProGluGlnLeuCysTrpIleSerGlyTrpGlyAlaThrGluGlu 389

Db 1151 GATGAGGAGCTCACTCCAGCCCACTCTGGATCATTTGGATGGGCTTTTACGAGCAG 1210

Qy 390 Lys---GlyLysThrSerGluValLeuAsnAlaAlaLysValLeuLeuIleGluThrGln 408

Db 1211 AATGGAGGGAAGATGTCTGACATCTGTCTGAGGCTCAGTCCAGTCTATTGACAGCACA 1270

Qy 409 ArgCysAsnSerArgTyrValTyrAspAsnLeuIleThrProAlaMetIleCysAlaGly 428

Db 1271 CGGTGCAATGCAAGCAGATCGTACCAGGGGAGGAGTCCAGAGAGATGATGTGCGAGC 1330

Qy 429 PheLeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyGlyProLeuValThrSer 448

Db 1331 ATCCCGAAGGGGTGTGGACACCTGCCAGGTGACAGTGGTGGGCCCTGATGTACCAA 1390

Qy 449 LysAsnAsnIleTrpTrpLeuIleGlyAspThrSerTrpGlySerGlyCysAlaLysAla 468

Db 1391 TCTGACAG---TGGCATGTGTGGGATCGTGTAGTGGGGCTATGGCTGGCGGGCCCG 1447

Qy 469 TyrArgProGlyValTyrGlyAsnValMetValPheThrAspTrpIleTyrArgGlnMet 488

Db 1448 AGACCCCAAGGATATACCAAGGTCTCAGGCTATCTCACTGGATCTCAATGTCTGG 1507

Qy 489 ArgAlaAsp 491

Db 1508 AAGGCTGAG 1516

RESULT 14

US-09-651-588-5

Sequence 5, Application US/09851588

Patent No. 6682890

GENERAL INFORMATION:

APPLICANT: Mack, David

APPLICANT: Gish, Kurt C.

APPLICANT: Wilson, Keith E.

TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND

TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS

FILE REFERENCE: A-68828-1/DJB/JDD/AMS

CURRENT APPLICATION NUMBER: US/09/851,588

CURRENT FILING DATE: 2001-09-24

PRIOR APPLICATION NUMBER: US 09/642,252

PRIOR FILING DATE: 2000-08-17

PRIOR APPLICATION NUMBER: US 09/656,002

PRIOR FILING DATE: 2000-09-06

NUMBER OF SEQ ID NOS: 9

SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 2079
TYPE: DNA
ORGANISM: Homo sapiens
US-09-851-588-5

Alignment Scores:
Pred. No.: 9,71e-58
Score: 687.50
Percent Similarity: 48.86%
Best Local Similarity: 35.40%
Query Match: 25.30%
DB: 4
Gaps: 18

US-09-615-285B-2 (1-492) x US-09-851-588-5 (1-2079)

Qy 41 ProAlaGlnTyrProSerProValProGlnTyrAlaProArgValLeuThr----- 58
Db 203 CCAGCATGGTAC-----AGGATCCTGCAGCATGATC 232
Qy 59 -----GlnAlaSerAsnProValValCysThrGlnProLysSerProSer 73
Db 233 AACCTCGAACAGCCTCGATGTCACACCC-----TGCAGAACCCCGTATCCCA--- 283
Qy 74 GlyThrValCysThrSerLysThrLysAlaLeuCysIleThrLeuThrLeuGlyThr 93
Db 284 TGGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCATCTACTGAGCCTGGCGAGT 343
Qy 94 PheLeuValGlyAlaAlaLeuAlaAlaGlyLeuLeuThrLysPheMetGlySerLysCys 113
Db 344 ATCATCATTTGGTGTTCCTCATCAAGTGATTCGATAAATCTAC----- 391
Qy 114 SerAsnSerGlyIleGluCysAspSerSerGlyThrCysIleAsnProSerAsnTrpCys 133
Db 392 -----TTCCCTCGGGCAGCCTCTCCACTTCATCCCGAGGAGCAGCTGTGT 439
Qy 134 AspGlyValSerHisCysProGlyGlyLysAspGlnAsnArgCysValArgLeuThr--- 152
Db 440 GACGAGAGCTGGAGCTGTCTCCCTTGGGGAGGACGAGGAGCAGCTGTGTCAAGAGCTTCCC 499
Qy 153 ---GlyPro-----AsnPheLeuLeuGlnValTrpSer 162
Db 500 GAAGGCTTCGAGTGGGAGTCCGGCTCTCCAGAGCCGATCCACACTGCAGGTCTGGAC 559
Qy 163 SerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGlnAsnTrpGlyArg 182
Db 560 TCGGCCACAGGAACTGTTCTCTGCTGTTTCGACAACTTCACAGAAAGCTCTCGCTGAG 619
Qy 183 AlaAlaCysArgAspMetGlyThr-----LysAsnAsnPheTrpSerSerGln----- 198
Db 620 ACAGCCTGTAGGCAGATGGGTACAGAGCAACCCACTTCAGAGCTGTGGAGATGGC 679
Qy 199 -----GlyIleValAspAspSerGlySerThrSerPheMetLysLeuAsn 213
Db 680 CCAGACAGGATCTGGATGTTGTAATCATCAGAAAACAGCCAGGAGCTTCGATCGG 739
Qy 214 ThrSerAlaGlyAsnValAspIleTrpLysLysLeuTrpHisSerAspAlaCysSerSer 233
Db 740 AACTCAAGTGGG-----CCCTGTCTCTCA 763
Qy 234 LysAlaValValSerLeuArgCysIleAlaCysGlyValAsnLeuAsnSerSerArgGln 253
Db 764 GGCTCCCTGGTCTCCCTGCAGCTGTCTGCTGTGGAGAGCCTG-----AAGACC 814
Qy 254 SerArgIleValGlyGlySerAlaLeuProGlyAlaTrpProTrpGlnValSerLeu 273
Db 815 CCCCGTGTGGTGGGAGGAGGAGGCTCTGTGGATTCCTGGCCCTTGGCAGGTGAGCATC 874
Qy 274 HisValGlnAsnValHisValCysGlySerIleThrProGluTrpIleValThr 293
Db 875 CAGTAGCAACACACACAGCTCTGTGGAGGAGCATCTCTGGACCCCTGAGTGGTCTCAGC 934

Qy 294 AlaAlaHisCysValGluLysProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGly 313
Db 935 CGAGCCCACTGCTTCAGGAACAT---ACGATGTGTCACTGGAAGTCCGGGAGGC 991
Qy 314 IleLeuArg---GlnSerPheMetPheTrpGlyAlaGlyTrpGlnValGluLysValIle 332
Db 992 TCAGACAAACTGGGAGCTTC-----CCATCCCTGGCTGTGGCCCAAGATCATC 1039
Qy 333 -----SerHisProAsnTrpAspSerLysThrLysAsnAsnAspIleAlaLeuMet 349
Db 1040 ATCATTTGAATTCACCCCATGTAC-----CCCAAGACATGACATCGCCTCATG 1090
Qy 350 LysLeuGlnLysProLeuThrPheAsnAspLeuValLysProValCysLeuProAsnPro 369
Db 1091 AAGCTGCAGTTCCTCCTCCTCAGGCACAGTCCAGGCCCATCTGTCTGCCCTTCTTT 1150
Qy 370 GlyMetMetLeuGlnProGluLeuLysTrpIleSerGlyTrpGlyValAlaThrGluGln 389
Db 1151 GATGAGGAGCTACTCCAGCCACCCCATCTGTGATCATTTGGGCTTTACGAAGCAG 1210
Qy 390 Lys---GlyLysThrSerGluValLeuAsnAlaLysValLeuLeuIleGluThrGln 408
Db 1211 AATGAGGAGGAGATGCTGTGACATACTGTCTGAGGCGTCCAGTCCAGTCAATTCACAGACA 1270
Qy 409 ArgCysAsnSerArgTrpValTrpAspAsnLeuLeuThrProAlaMetIleCysAlaGly 428
Db 1271 CGGTGCAATGACAGCATCGCTACCGGGGAGGATCCCGAGAGATGATGTGTCAGGC 1330
Qy 429 PheLeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyLysProLeuValThrSer 448
Db 1331 ATCCCGGAGGAGGTGTGACACACTGTCAGGCGTACAGTGTGGGCCCTGATGTACCAA 1390
Qy 449 LysAsnAsnIleTrpTrpLeuIleGlyAspThrSerTrpGlySerGlyCysAlaLysAla 468
Db 1391 TCTGACAG---TGGCATGTGTGGGATCGTTAGCTGGGCTATGGCTGGGGGGCCCG 1447
Qy 469 TyrArgProGlyValTrpGlyAsnValMetValPheThrAspTrpIleTrpArgGlnMet 488
Db 1448 AGCACCCAGGAGTATACACCAAGGTCTCAGCCTATCTCACTGGATCTACAATGTCTGG 1507
Qy 489 ArgAlaAsp 491
Db 1508 AAGGCTGAG 1516

RESULT 15

US-09-008-271A-18
Sequence 15, Application US/09008271A
Patent No. 6203979
GENERAL INFORMATION:

APPLICANT: Bandman, Olga
Hillman, Jennifer L.

Yue, Henry
Guegler, Karl J.

Corley, Neil C.

Tang, Tom Y.

Shah, Purvi

TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSES: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/008,271A

FILING DATE: 16-Jan-1998

PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PP-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLNNOT13
CLONE: 1337018
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-008-271A-18

Alignment Scores:
Pred. No.: 1.4e-56 Length: 2038
Score: 676.50 Matches: 166
Percent Similarity: 50.22% Conservative: 65
Best Local Similarity: 36.09% Mismatches: 161
Query Match: 24.90% Indels: 69
DB: 3 Gaps: 17

US-09-615-285B-2 (1-492) x US-09-008-271A-18 (1-2038)

QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 240 TCRAACCC-----TCCGCAACCCCGTATCCCATG-CAGACCTTC-----280
QY 81 ThrLysLysala-----LeuCysileThrLeuThrLeuGlyThrPheLeuVal 96
DB 281 ---AGAAAGGTGGGGATCCCATCATCATACATCATGAGCTGAGCTGGCGAGTATCATCAT 337
QY 97 GlyAlaAlaLeuAlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSer 116
DB 338 GTGGTTCTCTCATCAAGGTGATCTGGATAAATCTAC-----376
QY 117 GlyIleGluCysAspSerSerGlyThrCysileAsnProSerAsnTrpCysAspGlyVal 136
DB 377 ---TTCCTCTGGGGAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGAGAG 433
QY 137 SerHisCysProGlyGlyGluAspGluAsnArgCysValargLeuTyr-----GlyPro 154
DB 434 CTGGACTGTCCCTTGGGGGAGGACGAGGACACTGTGTCAAGAGCTTCCCGGAAGGGCCT 493
QY 155 -----AsnPheileLeuGlnValTyrSerSerGlnArg 165
DB 494 GCAGTGGCAGTCCGCCTCTCCAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACA 553
QY 166 LysSerTrpHisProValCysGlnAspTrpAsnGluAsnTyrGlyArgAlaAlaCys 185
DB 554 GGGAACTGGTCTCTGCTGCTGTTTCGACAACTTCACAGAGCTCTCGCTGACAGAGCCTGT 613
QY 186 ArgAspMetGlyTyr-----LysAsnAsnPheTyrSerSerGln-----198
DB 614 AGGCAGATGGGCTACAGCAACCAACCCACTTTAGAGCTGTGGAGATGGCCAGACCAG 673
QY 199 -----GlyIleValAspAspSerGlySerThrPheMetLysLeuAsnThrSerAla 216
DB 674 GATCTGGATGTTGTTGAATCACAGAAACAGCCAGGAGCTTCGCATGCGGAACCTCAAGT 733
QY 217 GlyAsnValAspIleTyrLysLeuTyrHisSerAspAlaCysSerSerLysAlaVal 236
DB 734 GGG-----CCCTGTCTCTCAGGCTCCCTG 757
QY 237 ValSerLeuArgCysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIle 256

758 GTCTCCCTGCACTGTCTTGGCTGTGGGAGAGCTG-----AAGACCCCGGTGTG 808
QY 257 ValGlyGlyGluSerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGln 276
DB 809 GTGGTGGGAGAGGCGCTCTGTGTGATTTCTTGGCTTGGCAGGTGACAGTACGAC 868
QY 277 AsnValHisValCysGlyGlySerIleLeuThrProGluTrpIleValThrAlaAlaHis 296
DB 869 AAACGACGAGCTGTGGAGGAGCATCTGACCCCACTGGGTCTCAGGAGCCAC 928
QY 297 CysValGluLysProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArg 316
DB 929 TGCTTCAGGAAACAT---ACCGATGTGTCACTGGAGGTGCGGCGAGGCTCAGACAAA 985
QY 317 ---GlnSerPheMetPheTyrGlyAlaGlyTyrGlnValGluLysValIle-----332
DB 986 CTGGGCACCTTC-----CCATCCCTGGCTGTGGCCAAGATCATCATATTCAA 1033
QY 333 SerHisProAsnTyrAspSerLysThrLysAsnAsnAspIleAlaLeuMetLysLeuGln 352
DB 1034 TTCRAACCCCATGTAC-----CCCAAGCAATGACATCGCCCTCATGAAGCTGAG 1084
QY 353 LysProLeuThrPheAsnAspLeuValLysProValCysLeuProAsnProGlyMetMet 372
DB 1085 TTCCCACTCATCTTCTCAGGCACAGTCAAGGCCCATCTGTCTGCCCTTCTTTGATGAGAG 1144
QY 373 LeuGlnProGluGlnLeuCysTrpIleSerGlyTrpGlyAlaThrGluGluLys---Gly 391
DB 1145 CTCACTCCAGCCACCCCACTCTGATCATCTGATGGGCTTTACGAGCAGAATGGAGG 1204
QY 392 LysThrSerGluValLeuAsnAlaAlaLysValLeuLeuIleGluThrGlnArgCysAsn 411
DB 1205 AAGATGTCTGACATACTCTGCAGCGCTCAGTCCAGGTTCATTGACAGCACACGCTGCAAT 1264
QY 412 SerArgTyrValTyrAspAsnLeuIleThrProAlaMetIleCysAlaGlyPheLeuGln 431
DB 1265 GCAGACGATGGTACCGGGGGAAGTCCACGAGAGATGATGTGTGAGGATCCCGGAA 1324
QY 432 GlyAsnValAspSerCysGlnGlyAspSerGlyGlyProLeuValThrSerLysAsnAsn 451
DB 1325 GGGGTGTGGACACCTGCCAGGTGACAGTGTGGGCCCTGTATCCAAATCTGACCAG 1384
QY 452 IleTrpTrpLeuIleGlyAspThrSerTrpGlySerGlyCysAlaLysAlaTyrArgPro 471
DB 1385 ---TGCCATGTGGTGGCATCGTTAGCTGGGGCTATGGCTCGGGGGCCCGACACCCCA 1441
QY 472 GlyValTyrGlyAsnValMetValPheThrAspTrpIleTyrArgGlnMetArgAlaAsp 491
DB 1442 GGAGTATACCAAGGTCTCAGCCTATCTCACTGATCTACATGTCTGGAAGCTGAG 1501

Search completed: June 1, 2004, 17:42:16
Job time : 159 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: June 1, 2004; 14:38:29 / Search time 629 seconds
(without alignments)
3322.915 Million cell updates/sec

Title: US-09-615-285B-2
Perfect score: 2717
Sequence: 1 MALNSGSPAPGYPYENHGY.....VGNVMVFTDWYQMRADG 492

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 337863 seqs, 2124099041 residues
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-MODEL=frame+ p2n.model -DB=xlh
-O=/cpn2.1/USPTO.spool/US09615285/runat 01062004 132132 19940/app query.fasta_1.647
-DB=N_Geneseq 237at04 -OPMT=fastap -SUPFLX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09615285 @CGN 1 1 470 @runat 01062004 132132 19940 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04: *
1: Geneseqn1980s: *
2: Geneseqn1990s: *
3: Geneseqn2000s: *
4: Geneseqn2001as: *
5: Geneseqn2001bs: *
6: Geneseqn2002s: *
7: Geneseqn2003as: *
8: Geneseqn2003bs: *
9: Geneseqn2003cs: *
10: Geneseqn2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2717	100.0	1479	3	Aaz87786 Human tum
2	2717	100.0	1738	3	Aaz29636 Human 20P
3	2717	100.0	1738	6	Aad28778 Human 20P
4	2712	99.8	2486	8	Adal0904 Human CDN
5	2712	99.8	2486	8	Acho0398 Human CDN
6	2711	99.8	3226	9	Adb75590 Prostate
7	2709	99.7	1476	4	Aah93944 P100C op
8	2709	99.7	1476	4	Aas64180 Human pro

9	2709	99.7	1476	5	ACA59988	ACA59988 Prostate
10	2709	99.7	1476	6	ABL95551	ABL95551 Human P10
11	2709	99.7	1476	7	ACC95715	ACC95715 Prostate
12	2709	99.7	1476	9	ADB14381	ADb14381 Human pro
13	2709	99.7	1479	4	Aah93943	Aah93943 P100C op
14	2709	99.7	1479	4	AAS64179	Aas64179 Human pro
15	2709	99.7	1479	5	ACA59987	ACA59987 Prostate
16	2709	99.7	1479	6	ABL95550	ABL95550 Human P10
17	2709	99.7	1479	7	ACC95714	ACC95714 Prostate
18	2709	99.7	1479	9	ADB14380	ADb14380 Human pro
19	2709	99.7	3245	3	AAA08803	AAA08803 Androgen-
20	2709	99.7	3245	4	Aah93942	Aah93942 P100C fu
21	2709	99.7	3245	4	AAS64178	Aas64178 Human pro
22	2709	99.7	3245	5	ACA59986	ACA59986 Prostate
23	2709	99.7	3245	6	ABL95549	ABL95549 Human P10
24	2709	99.7	3245	7	ACC95713	ACC95713 Prostate
25	2709	99.7	3245	9	ADB14379	ADb14379 Human pro
26	2709	99.7	3483	5	ABV21748	ABv21748 Human pro
27	2709	99.7	3483	5	ABV23318	ABv23318 Human pro
28	2709	99.7	3483	5	ABV24651	ABv24651 Human pro
29	2709	99.7	3483	5	ABV29119	ABv29119 Human pro
30	2709	99.7	3483	5	ABV27570	ABv27570 Human pro
31	2709	99.7	3483	5	ABV21767	ABv21767 Human pro
32	2709	99.7	3483	5	ABV29165	ABv29165 Human pro
33	2709	99.7	3483	5	ABV27589	ABv27589 Human pro
34	2704	99.5	3433	3	Aaz95005	Aaz95005 Cancer sp
35	2696	99.2	2479	3	Aaz87813	Aaz87813 Human tum
36	2696	99.2	2479	3	AAH93928	AAh93928 Human tra
37	2696	99.2	2479	4	AAS64164	Aas64164 Human CDN
38	2696	99.2	2479	4	AAD13168	Ad13168 Human ser
39	2696	99.2	2479	5	ACA59972	ACA59972 Prostate
40	2696	99.2	2479	5	ABL95535	ABL95535 Human tra
41	2696	99.2	2479	6	ABK32201	ABk32201 Prostate
42	2696	99.2	2479	6	AAD28779	Aad28779 Human TMP
43	2696	99.2	2479	7	ACC51058	Acc51058 Human bla
44	2696	99.2	2479	7	ACC95699	Acc95699 Human tra
45	2696	99.2	2479	7	ACC95699	Acc95699 Human tra

ALIGNMENTS

RESULT 1
AAZ87786
ID AAZ87786 standard; DNA; 1479 BP.
XX
AC AAZ87786;
DT 05-JUL-2001 (revised)
DT 12-MAY-2000 (first entry)
XX
DE Human tumour suppressor TMRSS2 encoding DNA.
XX
KW Tumour suppressor gene; TMRSS2; cancer; human; drug design;
KW gene therapy; protein therapy; ds.
XX
OS Homo sapiens.
XX
FH Key
CDS Location/Qualifiers
FT 1..1479
FT /tag= a
FT /product= "TMRSS2"
FT 478
FT /tag= b
FT /note= "this base can be G or A with G being the more
FT common allele. The codon will change from Val to Met."
FT 625
FT /tag= c
FT /note= "this base can be T or A with T being the more
FT common allele. The codon will change from Phe to Ile"
FT 724
FT conflict
FT /tag= d
FT /note= "listed as T in GenBank Accn No: U75329"
FT 768
FT allele

CC	20P1F12/TPMRSS2 gene (also designated 20P1F12-GTCL1, as deposited with ATCC accession number 207097). Anti-20P1F12/TPMRSS2 antibodies may be used as therapeutic agent for prostate and colon cancers, to image prostate cancer cells and prostate tumours, to identify ligands and cellular constituents that bind to a 20P1F12/TPMRSS2 gene product and for use as cancer vaccines
CC	Sequence 1738 BP; 433 A; 459 C; 473 G; 373 T; 0 U; 0 Other;
XX	
SX	
Alignment Scores:	
Pred. No.:	9,146-206
Score:	2717.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	3
US-09-615-285B-2 (1-492) x AAZ29636 (1-1738)	
QY	1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB	112 ATGGCTTTGAACCTAGGCTCACACAGCTATGGACCTTATATGAAAAACCATGATAC 171
QY	21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
DB	172 CAACCGGAAACCCCTATCCCGCACAGCCCACTGTGGTCCCCACTGTCTACGAGGTGCAT 231
QY	41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB	232 CCGGCTCTAGTACTACCGGTCCCGTCCCGAGTACGCCCGGAGGGTCTCTGACGCAAGCT 291
QY	61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB	292 TCCAAACCGGTCTGTGCAGCGAGCCCAATCCCCATCCGGGACAGTGTGCACCTCAAAG 351
QY	81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB	352 ACTAAGAAGCACTGTGCATCACTTGACCTCGGGAGACCTTCTCGTGGAGCTGCGCTG 411
QY	101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB	412 GCCGTGGCTACTCTGGAGTTCTATGGCAGCAAGTGTCTCAACTCTGGATAGAGTGC 471
QY	121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
DB	472 GACTCTCTCAGTACCTGCATCAACCCCTCTAACTGGTGTGATGGCGGTGTCACTGCC 531
QY	141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
DB	532 GCGCGGGAGGACGAACTGGTGTCTTCGCTCTACGGACCAAACTTCATCTTCAGGTG 591
QY	161 TyrSerSerGlnArgLysSerThrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
DB	592 TACTCATCTCAGAGAAGTCTCTGGCACCCCTGTGTGCCAAGCACTGGAACGAGAACTAC 651
QY	181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
DB	652 GCGCGGGCGGCTCTGAGGACATGGGCTATAGAATAATTTTCTCTAGCCAGGAATA 711
QY	201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB	712 GTGGATGACAGCGGATCCACCAGCTTTATGAACCTGAAACACAAGTCCGCGCAATGTCAT 771
QY	221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB	772 ATCTATAAAAACCTGACACAGTGTGCTCTTCTCAAAAGCAGTGGTTCCTTTACGC 831
QY	241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
DB	832 TGTATAGCTCGGGGTCACTTGAACTCAAGCCCGCAGAGCAGGATTTGTGGCGCGCAG 891
QY	261 SerAlaLeuProGlyAlaTyrProTyrProTyrProTyrProTyrProTyrProTyrProTyr 280

Db 892 AGCGCGCTCCGGGGGCTGGCTGGCAGGTGACGCTCCAGAACGTCACGCTG 951
Qy 281 CysGlySerLeuThrProGluThrAlaAlaHisCysValGluLys 300
Db 952 TGGGAGGCTCCATCATCACCCTGGAGTGGATGTCGACAGCCCGCTGGGAAAA 1011
Qy 301 ProLeuAsnProTTrpHisTTrpThrAlaPheAlaGlyLeuArgGlnSerPheMet 320
Db 1012 CCTCTTAACATCATGCGCATGGACGGCATTTCGGGGATTTTGACACATCTTTCA 1071
Qy 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 1072 TTTATGAGCGCGATACCAAGTAGAAAAAGTATTCTCATCAAAATTATGACTCCAG 1131
Qy 341 ThrLysAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1132 ACCAAGAACATGACATTCGCTGATGAAGCTGCAGAGGCTCTGACTTTCAACGACCTA 1191
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTTrp 380
Db 1192 GTGAACACGATGCTCTGCCCAACCCAGGATGATGCTGCAGCCAGAACAGCTCTGCTG 1251
Qy 381 IleSerGlyTTrpGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaLa 400
Db 1252 ATTTCCGGTGGGGGGCCACCGAGGAGAAAGGAGACCTCAGAGTGTGACGTGCC 1311
Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1312 AAGGTGCTTCTCATGACACACAGATGCAACAGACAGATGCTATGATGACACACCTGATC 1371
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnCysValAsnValAspSerCysGlnGlyAsp 440
Db 1372 ACACAGCCATGATCTGTGGCGGCTCTCTGCAGGGGAACGTGATTTCTCCAGGGTGAC 1431
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTTrpLeuIleGlyAspThrSer 460
Db 1432 AGTGGAGGCTCTGTGCTTCCGAGAACATATCTGTGGTGTAGGGGATACCAAGC 1491
Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1492 TGGGGTTCTGGCTGTGCCAAAGCTTACAGACACAGAGTGTACGGGATGTGATGGTATTC 1551
Qy 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1552 ACGAGTGGATTATTCGACAAATGAGGGCAGACGGC 1587

RESULT 3

AAD28778 standard; cDNA; 1738 BP.

XX AAD28778;

XX 07-MAY-2002 (first entry)

XX Human 20P1F12-GTC1 cDNA.

XX Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTC1; cell growth; neoplasm;
XX cancer; vaccine; human; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
XX 112..1591
XX CDS /*tag= a

XX /product= "Human 20P1F12-GTC1 protein"

XX WO200204953-A2.

XX 17-JAN-2002.

XX 12-JUL-2001; 2001WO-US022168.

XX 12-JUL-2000; 2000US-00615285.

XX (AGEN-) AGENSYS INC.
XX Saferran D, Raitano AB, Hubert RS, Jakobovits A, Faris M;
XX Challita-Bid PM;
XX WPI; 2002-154967/20.
XX P-PSDB; AAE18096.
XX Examining a biological sample for evidence of dysregulated cellular
XX growth, comprises comparing the status of prostate-specific, androgen-
XX regulated, secreted serine protease, 20P1F12/TMPRSS2, in a corresponding
XX normal sample.
XX Example 3; Fig 1; 161pp; English.
XX The present invention relates to methods and compositions for the
XX diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
XX kidney cancer derived from or based on a normally prostate-specific,
XX androgen regulated, cell membrane associated secreted serine protease
XX termed 20P1F12/TMPRSS2. The invention further relates to a method of
XX examining a biological sample for evidence of dysregulated cellular
XX growth comprises comparing the status of 20P1F12/TMPRSS2 gene (also
XX designated 20P1F12-GTC1) in the sample to the status of 20P1F12/
XX in a corresponding normal sample. The invention also relates to 20P1F12/
XX TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
XX invention are used for examining a sample such as blood, serum, stool,
XX urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
XX The dysregulated cell growth is indicative of bladder cancer, lung
XX cancer, kidney cancer or ovarian cancer. It is useful for identifying
XX evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
XX portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
XX expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
XX lung cancer, ovarian cancer or metastatic cancer. The present sequence is
XX human 20P1F12-GTC1 cDNA
XX Sequence 1738 BP; 433 A; 459 C; 473 G; 373 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 9, 14e-206 Length: 1738
Score: 2717.00 Matches: 492
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-615-285B-2 (1-492) x AAD28778 (1-1738)

Qy 1 MetaLalaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 112 ATGGCTTTGAATCTAGAGGTCTACCCAGCTATTGGACCTTACTATGAAACCATGGATAC 171
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 172 CAACCGGAACCCCTATCCGACAGCCCACTGTGGTCCCACTGTCTACGAGTGCAT 231
Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 232 CCGGCTCAGTACTACCCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCTCC 291
Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 292 TCCAAACCCCGCTGTGTGACGAGCCCAATCCCAATCCCAATCCCAATCCCAATCCCA 351
Qy 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
Db 352 ACTAAGAAGACACTGTGCATCACCTTGACCTCCCTGGGAGCCTTCTCTGGGAGCTCGCTG 411
Qy 101 AlaAlaGlyLeuLeuTTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 412 GCCGCTGGCCTACTCTGGAAGTTTCATGGCAGCAAGTGTCCAAATCTGGGATAGATGC 471
Qy 121 AspSerSerGlyThrCysIleAsnProSerAsnTTrpCysAspGlyValSerHisCysPro 140

```

Db 472 GACTCTCAGTACTGCTGATCAACCCCTCTAACTGGTGTGATGCGGTGCACACTGCCCC 531
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 532 GCGCGGGAGGACGAGAAATCGGTGTGTTCTGCTCTACGGACCAAACTTCATCCTTCAGGTG 591
Qy 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Db 592 TACTCATCTCAGAGGAAGTCTGGCACCTGTGTGCCAAGACGACTGGACGAGAACTAC 651
Qy 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
Db 652 GCGCGGGCGGCTCGAGGACATGGCTATAGGAATAATTTTACTCTAGCAAGGAATA 711
Qy 201 ValAspSerSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 712 GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACACAGTGGCGGCAATGTGAT 771
Qy 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 772 ATCTATAAAACCTGTACACAGTGATGCCCTGTCTTCAAAGCAGTGGTTTCTTTACGC 831
Qy 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 832 TGTATAGCTCGGGGGTCACTTGACTCAAGCCGCCAGACGAGATTTGGGCGGCGAG 891
Qy 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 892 AGCGCGCTCCCGGGGCGCTGGCCCTGGCAGTCCAGCTGACGCTCCAGACCTCCAGTG 951
Qy 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db 952 TCGGAGAGCTCCATCATCACCCCGAGTGGATCGTACAGCGGCCCTGCGTGGGAAAA 1011
Qy 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 1012 CCTCTTACATCCATGCTGACGCGCATTTGGCGGATTTTGACACAATCTTTCATG 1071
Qy 321 PheTyrGlyAlaGlyTyrGlnValGlyLysValIleSerHisProAsnTyrAspSerLys 340
Db 1072 TTCTATGAGCGCGGTACCAAGTAAAAAGTATTTCTCATCCAAATTTATGATCCCAAG 1131
Qy 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1132 ACCAAGAACATGCAATTCGCGCTGATGAAGCTGCAGAACCTCTGACTTTCACAGCCTA 1191
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1192 GTGAACCCAGTGTGTCTGCCAACCCAGCGCATGCTGCAGCCAGAACAGCTCTGCTGG 1251
Qy 381 IleSerGlyTrpGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1252 ATTTCCGGGTGGGGGGCCACCGAGAGAAAGGAGACCTCAGAAAGTCTGACAGCTGCC 1311
Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1312 AAGTGTCTTCTCATTGAGACACAGAGATGCAACAGCAGATATGTCTATGACAACTGATC 1371
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1372 ACACAGCCATGATCTGTGCGGCTCTCTGAGGGGAACGTCGATTTCTGCGAGGGTGAC 1431
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1432 AGTGAGGGCCTCTGGTCACTTCGAGAACCAATATCTGTGTGCTCATAGGGGATACAGC 1491
Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1492 TGGGGTCTCTGGTGTGCCAAAGCTTACAGACAGGAGTGTACGGGAATGTGATGTTATC 1551
Qy 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492

```

```

Db 1552 ACGGACTGGATTATCGACAAATGAGGGCAGACGGC 1587
RESULT 4
ADAI0904
ID ADA10904 standard; cDNA; 2486 BP.
XX AC ADA10904;
XX DT 06-NOV-2003 (first entry)
XX DE Human cDNA differentially expressed in colon cancer #15.
XX ES ss; differential expression; colon cancer; cancer; human; gene.
XX OS Homo sapiens.
XX PN US2002160382-A1.
XX PD 31-OCT-2002.
XX PF 11-OCT-2001; 2001US-00981353.
XX PR 11-OCT-2000; 2000US-0239841P.
XX PA (LASE/) LASEK A W.
XX PA (JONE/) JONES D A.
XX PI Lasek AW, Jones DA;
XX DR WPI; 2003-265756/26.
XX DR P-PSDB; ADA10905.
XX PT New combination comprising cDNAs that are differentially expressed in
XX PT colon disorder, useful for diagnosing, treating, staging or monitoring
XX PS treatment for colon cancers.
XX PS Claim 1; SEQ ID NO 22; 231bp; English.
XX CC The invention relates to a combination comprising cDNAs that are
XX CC differentially expressed in colon disorder. The methods and compositions
XX CC of the present invention are useful for diagnosing, treating, staging or
XX CC monitoring treatment for colon cancer. They are also useful in high
XX CC throughput methods for using cDNAs to detect differential expression of
XX CC nucleic acids in a sample, screening molecules or compounds to identify a
XX CC ligand which specifically binds a cDNA and using a protein to screen
XX CC molecules or compounds to identify at least one ligand which specifically
XX CC binds the protein. The present sequence represents a human cDNA
XX CC differentially expressed in colon cancer.
XX SQ Sequence 2486 BP; 564 A; 660 C; 719 G; 543 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,58e-205 Length: 2486
Score: 2712.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.82% Indels: 0
DB: Gaps: 0

US-09-615-285B-2 (1-492) x ADA10904 (1-2486)
Qy 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 149 ATGGCTTTGAACCTCAGGGTCACCCCTGCTATTTGGACCTTACTATGAAAACCAATGGATAC 208
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
Db 209 CAACCGGAAAACCCCTATCCGCACAGCCACTGTGGTCCCACTGTCTACGAGGTGCAT 268
Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 269 CCGGCTCAGTACTACCCGTCGCCCGTCCCGGTCAGTACGCCCGGAGGTCTCTGACGCGGCT 328

```

QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 329 TCCAAACCCCGTCTGCTGACGACGACCAATATCCCATCCGGACAGTGTGCACCTCAAAG 388
QY 81 ThrLysLysAlaLeuCysLysLeuThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
Db 389 ACTAAGAAAGCACTGTGCATCCTTGACCTCGGGACCTTCTCTGCGGAGCTGCGCTG 448
QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyLysLeuCys 120
Db 449 GCCCTGGCTACTCTGGAGTTCATGGGACGACGAGTCTCCAACTCTGGGAGTAGTGC 508
QY 121 AspSerSerGlyThrCysLysLeuAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
Db 509 GACTCTCTCAGTACCTGCATCAACCCCTTAACCTGGTGTGATGGCGTGTCCACTGCCCC 568
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheLeuGlnVal 160
Db 569 GCGGGGAGGACGAGATCGGTGTGCTGCTCTACGACCAAACTTCATCTCTCAGGTG 628
QY 161 TyrSerSerGlnArgLysSerTyrPheProValCysGlnAspAspTyrAsnGluAsnTyr 180
Db 629 TACTCATCTCAGAGAAAGTCTGCGACCTGTGTGCCAAGACGACTGGAAACGAGAACTAC 688
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLys 200
Db 689 GCGCGGGCGGCTCAGAGACATCGGTGTGCTCTACGACCAAACTTCATCTCTCAGGTG 748
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 749 GTGGATGACAGCGGATCACCGCTTTATGAACCTGAACCAAGTGGCGGCAATTCGAT 808
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValSerLeuArg 240
Db 809 ATCTATATAAAACGTACACAGATGCTGCTGTCTTCAAAGCAGTGGTTCCTTACGC 868
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
Db 869 TGTATAGCTCGGGGTCACTTGAACCTCAAGCCGACAGAGAGTGTGGCGCGAG 928
QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
Db 929 AGCGCTCCCGGGGCGCTGCGCTGCGAGCTGAGCTGCGACGCTCCAGAAACGTCACAGT 988
QY 281 CysGlyGlySerIleLeuThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
Db 989 TCGGAGGCTCCATCATCACCCCGAGTGGACCGTGCAGCCGCCACTCGGTGGAAAAA 1048
QY 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 1049 CCTCTTAACAATCCATGCAATTTGACGGCAATTGCGGGGATTTGAGACAACTTTTCATG 1108
QY 321 PheTyrGlyAlaGlyTyrGlnValLysValIleSerHisProAsnTyrAspSerLys 340
Db 1109 TTCTATGGAGCGGATACCAAGTGAAGAAAGTATTTCTCATCCAAATTTAGCTCCAG 1168
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1169 ACCAAGACATGACATTCGCTGATGAAGCTGAGAGCTCTGACTTCAACGACCTA 1228
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLysCysTyr 380
Db 1229 GTGAACACAGTGTGTCTCCCAACCCAGGATGATGTCTGACGACGACAGCTCTGTCTG 1288
QY 381 IleSerGlyTyrGlyValAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1289 ATTTCGGGTGGGGGCGCCAGGAGGAGGAGGAGACCTCAGAGTGTCTGAACGCTGCC 1348
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1349 AAGGTGCTTCTCATTTGACACACAGAGATGCAACAGCAGATATGTCTATGACAACTGATC 1408
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440

Db 1409 ACACAGCCATGATCTGTGCGGCTTCTGAGGGGAACGTCGATTCTTGCCAGGGTGAC 1468
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
Db 1469 AGTGAGGGCTCTGCTCCTTCGAGAACCAATATCTGCTGCTGATAGGGGATACAAGC 1528
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1529 TGGGGTCTGCTGCTGCCAAAGCTTACAGACCAAGAGTGTACGGGAATGTGTGTATTC 1588
QY 481 ThrAspTyrPheTyrArgGlnMetArgAlaAspGly 492
Db 1589 ACGGACTGATTTATCGACAAATGAGGCGACAGCGC 1624
RESULT 5
ACH03998
ID ACH03998 standard; cDNA; 2486 BP.
XX ACH03998;
AC ACH03998;
XX 26-SEP-2003 (first entry)
DT Human cDNA differentially expressed in lung cancer #203.
DE Gene therapy; emphysema; ss; gene; chronic obstructive pulmonary disease;
KW respiratory disorder; lung cancer; asthma; human.
KW
XX Homo sapiens.
OS
XX US2003065157-A1.
XX
XX 03-APR-2003.
XX
XX 04-APR-2002; 2002US-00116802.
XX
XX 04-APR-2001; 2001US-0281593P.
XX
XX (LASE/) LASEK A W.
XX
XX Lasek AW;
XX
XX WPI; 2003-540803/51.
XX
XX New combination comprising cDNAs that are differentially expressed in
PT respiratory disorders, useful for diagnosing or treating respiratory
PT disorders e.g., lung cancer, chronic obstructive pulmonary disease,
PT emphysema or asthma.
XX
XX Claim 1; Page; 39pp; English.
XX
XX The invention relates to a combination comprising cDNAs or their
CC complements that are differentially expressed in respiratory disorder.
CC The combination is useful for preparing a composition for diagnosing or
CC treating respiratory disorders e.g. lung cancer, chronic obstructive
CC pulmonary disease, emphysema or asthma. The present sequence represents
CC human cDNA differentially expressed during lung cancer
XX
SQ Sequence 2486 BP; 564 A; 660 C; 719 G; 543 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 3 58e-205 Length: 2486
Score: 2712.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.82% Indels: 0
DB: Gaps: 0
US-09-615-285B-2 (1-492) x ACH03998 (1-2486)
QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 149 ATGGCTTTGAACTCAGGTCACCACTGCTATTTGACCTTACTATGAAACCAACCATGATAC 208

CC ADB75177-ADB75631 represent marker cDNA and proteins. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences.

XX SQ Sequence 3226 BP; 782 A; 808 C; 857 G; 779 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 5, 96e-205 Length: 3226
Score: 2711.00 Matches: 490
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 99.78% Indels: 0
DB: 9 Gaps: 0

US-09-615-285B-2 (1-492) x ADB75590 (1-3226)

QY	1	MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr	20
DB	128	ATGGCTTTGAACCTCAGGCTCACCACCACTATTGGACCTTACTATGAAACCACTGGATAC	187
QY	21	GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis	40
DB	188	CAACCGGAACCCCTATCCCGCACAGCCCACTGTGTCCTCACTGTCTACGAGGTGCAT	247
QY	41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60
DB	248	CCGGCTCAGTACTACCCCTCCCGCTCCCGCAGTACCGCCGAGGGTCCCTGACGACGGCT	307
QY	61	SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys	80
DB	308	TCCAAACCCCGTCTGCACAGCCGCCAATCCCATCCCGGACAGTGTGCACCTCAAG	367
QY	81	ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu	100
DB	368	ACTAAGAAAGCACTGTGCATCACCTTGACCTGGGGACCTTCTCTCGTGGGAGCTGGCTG	427
QY	101	AlaAlaGlyLeuLeuTyrPhePheGlySerLysCysSerAsnSerGlyIleGluCys	120
DB	428	CCCGCTGGCCCTACTCTGGAAGTTCATGGCGAGAGTGTCTCCAACTCTGGATAGAGTGC	487
QY	121	AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro	140
DB	488	CACCTCTCAGGTACCTGCATCAACCCCTCTAACTGGTGTGATGGCTGTGCACACTGCC	547
QY	141	GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal	160
DB	548	GGCGGGAGGACGAGATCGGTGTGTTCGCCCTCTACGACCAAACTTCATCTCTTCAGATG	607
QY	161	TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr	180
DB	608	TACTCATCTCAGAGGAAGTCTGGCACCCTGTGTGCCAAGACGACTGGACGAGACTAC	667
QY	181	GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle	200
DB	668	GGCGGGCGGCTGCAAGGACATGGGCTATAGAAATAATTTTACTCTAGCCAAAGGAATA	727
QY	201	ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp	220
DB	728	GTGGATGACAGCGGATCCACAGCTTTATGAACTGAACAAAGTGGCGCAATGTGCAT	787
QY	221	IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg	240
DB	788	ATCTATAAAAAAATGTACACAGTGTATGCCCTGTCTTCAAAAGCAGTGGTTTCTTTACGC	847
QY	241	CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu	260
DB	848	TGTATAGCTCGGGGGTCACTTTGAACTCAAGCCGCCAGACGAGATTTGGGGCGGTGAG	907
QY	261	SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal	280
DB	908	AGCGCGCTCCCGGGGCGCTGGCCCTGGCAGGTTCAGCTGCAGCTCCAGAACCTCCACGTG	967

QY	281	CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys	300
DB	968	TGCGAGAGCTTCATCATCACCCCCGAGTGGATCGTGACAGCGCCCACTGGTGAAAAA	1027
QY	301	ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet	320
DB	1028	CCTCTTAACAATCCATGGCATTTGGACGGCATTTGGGGGATTTTGAGCAATCTTTTCATG	1087
QY	321	PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys	340
DB	1088	TTCTATGGAGCCGGATACCAAGTAGAAAAGTGAATTTCTCATCAAAATATGACTCCAAG	1147
QY	341	ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu	360
DB	1148	ACCAAGAACAATGACATTCGCTGATGAAGCTGCAAGAGCTCTGACTTTCAACGACCTA	1207
QY	361	ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp	380
DB	1208	GTGAACACAGTGTGTCTGCCCAACCCAGGCATGCTGCAGCCAGAACAGCTCTGCTGG	1267
QY	381	IleSerGlyTyrGlyAlaThrGluGlyLysThrSerGluValLeuAsnAlaAla	400
DB	1268	ATTTCGGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAAAGTGTGAAACGCTGCC	1327
QY	401	LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle	420
DB	1328	AAAGTGTCTCTCTTGTAGACACAGAGATGCACACAGATATGTCTATGACACCTGATC	1387
QY	421	ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp	440
DB	1388	ACACCCAGCCATGATCTGTGCCGGCTTCTGCAGGGGAAACGTGATTTCTTGCAGGGTGAC	1447
QY	441	SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer	460
DB	1448	ATGGAGGGCCCTCTGGTCACTTCGAGAAACAATATCTGGTGGCTGATAGGGGATACAGC	1507
QY	461	TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe	480
DB	1508	TGGGGTCTCGCTGTGCCAAAGCTTACAGACCAGAGTGTACGGGAATGTGATGTATTTC	1567
QY	481	ThrAspTrpIleTyrArgGlnMetArgAlaAspGly	492
DB	1568	ACGGACTGGATTTATCGACAAATAGGGCAGACGGC	1603

RESULT 7

AAH93944
ID AAH93944 standard; cDNA; 1476 BP.
XX AC AAH93944;
XX DT 04-OCT-2001 (first entry)
XX DE P1000C open reading frame cDNA sequence without stop codon.
XX KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
XX KW cytostatic; gene therapy; metastasis; ss.
XX OS Homo sapiens.
XX PN WC200151633-A2.
XX PD 19-JUL-2001.
XX PF 16-JAN-2001; 2001WO-US0001574.
XX PR 14-JAN-2000; 2000US-00483672.
XX PA (CORI-) CORIXA CORP.
XX PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;
XX

DR WPI; 2001-425873/45.
 XX New polynucleotide encoding a prostate-specific protein, for diagnosing,
 PT monitoring and treating prostate cancer in a patient and for use in
 PT vaccines.
 XX
 XX Claim 1; Page 538-539; 543pp; English.
 XX
 XX The present invention describes polynucleotide sequences (I) which encode
 CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
 CC and can be used in vaccine production and gene therapy. (I), (II),
 CC antibodies to (I), fusion proteins comprising (II), and isolated T cells
 CC prepared using (I) or (II) are used treat cancer in a patient. (I) and
 CC the antibodies are also used in the detection of cancer in a patient. The
 CC cancer that is diagnosed or treated is particularly prostate cancer. (I)
 CC and (II) can be used in vaccines. The antibodies or (I) can be used for
 CC monitoring the progression of cancer in a patient. (I) and (II) can also
 CC be used to improve diagnostic and therapeutic methods for prostate
 CC cancer. They can indicate the level of metastasis as well as the prostate
 CC volume. AAH93357 to AAH93944 and AAH0115 to AAH01318 represent
 CC polynucleotide and amino acid sequences used in the exemplification of
 CC the present invention
 XX
 SQ Sequence 1476 BP; 361 A; 405 C; 402 G; 308 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3, 21e-205 Length: 1476
 Score: 2709.00 Matches: 491
 Percent Similarity: 99.80% Conservativity: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 99.71% Indels: 0
 Gaps: 0

US-09-615-285B-2 (1-492) x AAH93944 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Db 1 ATGGCTTTGAACCTCAGGCTCACCACAGCTATTGGACCTTACTATGAACCAATGGATAC 60
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
 Db 61 CAACCGGAAACCCCTATCCCGCAGACGCCACTGTGGTCCCACTGTCTCAGAGGTGCAT 120
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 121 CGGGCTCAGTACTACCCGTCCTCCCGCCAGTACGACCCCGGAGGCTCTCAACGAGCT 180
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 181 TCCAAACCCCGTCTGTGCAGCGAGCCCAATCCCAATCCCGACAGTGTGCACCTCAAG 240
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 Db 241 ACTAAGAAAGCACTGTGCATCACTTGCCTGGGAGCTTCTCTGTGGAGCTGGCGTG 300
 QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 301 GCGCGTGGCTACTCTGGAAGTTTCATGGCAGCAAGTGTCTCAACTCTGGGATAGAGTGC 360
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 Db 361 GACTCTCAGGTACTGTGCATCAACCCCTCTAACTGTGTGTGGGTGTGCACCTGCCCC 420
 QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 421 GCGCGGAGGACGAGAAATCGGTGTGTTCGCTCTACGGATCAAACTTCACTCTCAGGTG 480
 QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAsnTyrAsnGluAsnTyr 180
 Db 481 TACTCATCTCAGAGAAGTCTCGGACCCCTGTGTGCCAAGACGACTGGAGCAGACTAC 540
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlyGlyIle 200

Db 541 GGCGCGCGCGCTGCAGGACATGGGCTATAGAAATAATTTTACTCTAGCCAGGAATA 600
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 601 GTGGATGACAGCGGATCCACCGAGCTTATGAAACTGAACACCAAGTCCCGCAATGTGAT 660
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Db 661 ATCTATATAAAATCTGTACCACAGTGTATGCTGTCTTCAAAAGCAGTGTGTCTTTACGC 720
 QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 Db 721 TGTATAGCTCGCGGCTCAACTTGAACCTCAAGCCGCGCAGCAGGATTTGTGGCGCGAG 780
 QY 261 SerAlaLeuProGlyAlaIleProProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 Db 781 AGCGCGCTCCCGGGGCGCTGGCCCTGGCAGGTGAGCTCCAGCTCCAGAACGTCACAGTG 840
 QY 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
 Db 841 TGGGAGGCTCCATCATCAACCCGAGTGTGATCGTACAGCCGCCCTGCTGGTGGAAAAA 900
 QY 301 ProLeuAsnAsnProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 Db 901 CCTCTTAACAATCCATGGCATTTGACGGCATTTGCGGGATTTTGAGACCAATCTTTCATG 960
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 Db 961 TTCTATGGAGCGGATACCAAGTAGAAAAAGTAGTTTCTATCCAAATATATGACTCCAG 1020
 QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 Db 1021 ACCAAGAACAATGACATTTGGCTGTGTAAGTGTGAGAGCCTCTGACTTTCAACGACCTA 1080
 QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGlnGlnLeuCysTrp 380
 Db 1081 GTGAAACCAAGTGTGTCTGCCCAACCCAGGATGTATCTGCAGCAGACAGCTCTCTGG 1140
 QY 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 Db 1141 ATTTCCGGTGGGGGCGCCACCGAGGAGAAAGGAGACCTCAGAGTGTGTGAACGCTGCC 1200
 QY 401 LysValLeuLeuIleGlnThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 Db 1201 AAGGTCTTCTCATTTAGACACAGATGCAACAGCAGATATGTCTATGACAACTGTATC 1260
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 Db 1261 ACACCGACCATGTATCTGTCCGCTTCTGTGAGGGAACGTGATTTCTTGGCAGGCTGAC 1320
 QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
 Db 1321 AGTGGAGGCGCTCTGTCTCACTTCGAAGAACAATATCTGTGTGGCTGATAGGGGATACA 1380
 QY 461 TrpGlySerCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 Db 1381 TGGGGTCTGGCTGTGCCAAAGCTTACAGACCGAGAGTGTACGGGATGTGATGTATTC 1440
 QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
 Db 1441 ACGGACTGGATTTATCGACAAATGAGGGCAGACCGC 1476

RESULT 8
 AAS64180
 ID AAS64180 standard; cDNA; 1476 BP.
 XX
 AC AAS64180;
 XX
 DT 29-JAN-2002 (first entry)
 XX
 DE Human prostate cDNA sequence #599.
 XX
 KW Human; prostate cancer; ss; cytostatic; immunostimulant; tumour.

XX OS Homo sapiens.
XX PN W0200173032-A2.
XX PD 04-OCT-2001.
XX PF 27-MAR-2001; 2001WO-US009919.
XX PR 27-MAR-2000; 2000US-00536857.
XX PR 09-MAY-2000; 2000US-00568100.
XX PR 12-MAY-2000; 2000US-00570737.
XX PR 13-JUN-2000; 2000US-00593793.
XX PR 27-JUN-2000; 2000US-00605783.
XX PR 09-AUG-2000; 2000US-00636215.
XX PR 29-AUG-2000; 2000US-00651236.
XX PR 06-SEP-2000; 2000US-00657279.
XX PR 02-OCT-2000; 2000US-00679426.
XX PR 10-OCT-2000; 2000US-00685166.
XX PR 09-NOV-2000; 2000US-00709729.
XX PF (CORI-) CORIXA CORP.
XX PI Xu J, Dillon DC, Mitcham JL, Harlocker SH, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX PF WPI: 2001-539232/73.
XX DR F-PSDB; AAU69960.
XX PT New human prostate-specific polypeptides and polynucleotides useful for
XX the diagnosis and treatment of cancer, especially prostate cancer.
XX PS Claim 1; Page 572; 579pp; English.
XX CC The invention relates to isolated prostate-specific polynucleotides,
XX polypeptides, fusion proteins of the polypeptides, antibodies raised
XX against the polypeptides (or antigenic epitopes derived from them) and
XX antigen-presenting cells expressing the polypeptides. The antibodies are
XX useful for detecting the presence of cancer, especially prostate cancer.
XX CC The polypeptides, polynucleotides and the antigen-presenting cells are
XX useful for stimulating and/or expanding T cells specific for a tumour
XX protein, and for inhibiting the development of cancer especially prostate
XX cancer. Compositions comprising the polynucleotide and/or polypeptide are
XX useful for stimulating an immune response, and for treating cancer. The
XX oligonucleotide is useful for detecting cancer. The present sequence is a
XX prostate specific polynucleotide of the invention
SQ Sequence 1476 BP; 361 A; 405 C; 402 G; 308 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,21e-205 Length: 1476
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 4 Gaps: 0

US-09-615-285B-2 (1-492) x AAS64180 (1-1476)

QY 1 MetAlaLeuAenSerGlySerProProAlaileGlyProTyrTyrGluAenHisGlyTyr 20
DB 1 ATGGCTTTGAACCTCAGGCTCACCACAGCTATTGGACCTTACTATGAAACCAACCATGGATAC 60
QY 21 GlnProGluAenProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
DB 61 CAACCGGAAACCCCTATCCCGCAGCCACTGTGGTCCCACTGCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CCGGCTCAGTACTACCCCTCCCGGTGCCAGTACGCCCGCGAGGTCCTGCAGCAGGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80

DB 181 TCCAAACCCCGTCTGCACGACGCCCAANTCCCACTCCGGGACAGTGTGACCTCAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
DB 241 ACTAAGAAAGCACTGTGTGCATCACTTGAACCTGGGACCTTCTCTCGTGGAGTGCCTG 300
QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerLysSerGlyIleGluCys 120
DB 301 GCCGCTGGCCCTACTCTGGAGTTCTGGGACGAGTGTCTCCAACTCTGGGATAGAGTGC 360
QY 121 AppSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
DB 361 GACTCTCAGGTACTGTGCATCAACCCCTCTAACTGGTGTGATGGCGTGTCCACACTGCC 420
QY 141 GlyGlyGluAspGluAenArgCysValArgLeuTyrGlyProAenPheIleLeuGlnVal 160
DB 421 GCGGGGAGGACGAGATCGGTGTGGCTCTACGATCAAACTTCATCTTCAGGTG 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
DB 481 TACTCATCTCAGAGGAAGTCTCGCACCTGTGTGCCAAGACGACGTGGAACGAGAATAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAenAsnPheTyrSerSerGlnGlyIle 200
DB 541 GCGCGGGCGGCTCGCAGGACATGGCTATAAAGATATAATTTTACTCTAGCAAGGAATA 600
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB 601 GTGGATGACACGGGATCCACAGCTTTATGAACTGAACCGCCAGACAGTCCCGCAATGCGAT 660
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 ATCTATAAAACCTGTACCACTGATGCTCTTCAAAAGCAGTGGTTCCTTTACGC 720
QY 241 CysIleAlaCysGlyValAenLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
DB 721 TGTATAGCTCGGGGCTCAACTTGAACCTCAAGCGCCAGACAGGATTTGGCGCGGAG 780
QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAenValHisVal 280
DB 781 AGCGGCTCCCGGGGCTGCGCTGCGAGTCACTGACGCTCCAGAACGTCACGCTG 840
QY 281 CysGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
DB 841 TCGGAGGCTCCATCATCACCCTGAGTGGATCGTACAGCCGCCCACTGCTGGGAAAA 900
QY 301 ProLeuAenAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
DB 901 CCTCTTAACAATCCATGGCATTGGACGGCATTTTCGGGGATTTTGAGACAATCTTTCA 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGlyValIleSerHisProAsnTyrAspSerLys 340
DB 961 TTTATGGAGCGGATACCAAGTAGAAAAAGTATTTCTCATCAAATATATGATCTCAAG 1020
QY 341 ThrLysAenAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAspLeu 360
DB 1021 ACCAAGAACAATGACATTTGGCTGATGAGCTCCAGAGCTCTGACTTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
DB 1081 GTGAACACAGTGTGTCTGCCCAACCCAGGCATGATCTCCAGCCAGCAACAGCTCTCTGG 1140
QY 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAenAlaAla 400
DB 1141 ATTTCCGGTGGGGGCCCGCAGGAGAAAGGAAGACCTTCAGAGTGTGTGAACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAenSerArgTyrValTyrAspAenLeuIle 420
DB 1201 AAGTGTCTTCTCATTTAGACACAGATGCAACAGCAGATATGTCTATGACACCTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440


```

QY 221 IletyrlslyslsleuTyHrHsSerAspAlaCySerSerlyeAlaValValSerLeuArg 240
DB 661 ATCTATAAAACCTGTACCAACAGTATGCTCTTCTTCAAAGCAGTGGTTCTTTACGC 720

QY 241 CysileAlaCysGlyValAenLeuAenSerSerArglnSerArgileValGlyGlyGlu 260
DB 721 TGTATAGCCTGCGGGGTCAACTTGAATCAAGCCGCGAGCAGGAGTTGTGGGCGCGAG 780

QY 261 SerAlaLeuProGlyAlaTtrPTrGlnValSerLeuHisValGlnAenValHisVal 280
DB 781 AGCGCGCTCCGGGGGCTGGCCCTGGCAGCTGAGCTGCAGCTCCAGAACGTCACAGTG 840

QY 281 CysGlySerleuThrProGluTtrPTrIleValThrAlaAlaHisCysValGlyLys 300
DB 841 TGGGAGGCTCCATCATCAACCCGAGTGATGCTGACAGCCGCGCTGCGTGGAAAAA 900

QY 301 ProLeuAenAenProTtrHsTtrAlaPheAlaGlyIleLeuArglnSerPheMet 320
DB 901 CCTCTTAACAAATCATGGCATTTGACGCGATTTGCGGGATTTTGAGACAACTTTTCAIG 960

QY 321 PheTyrglyAlaGlyTyrglnValGlyValIleSerHisProAenTyrrAspSerLys 340
DB 961 TTCTATGAGCCGATACCAAGTAGAAAAGTGTATTTCTCATCAAAATTATGACTCCAAG 1020

QY 341 ThrlyAenAenAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAspLeu 360
DB 1021 ACCAAGAACAAATGATGCTGATGAAGCTGACAGAGCCTCTGACTTTCAACGACCTA 1080

QY 361 ValLysProValCysLeuProAenProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
DB 1081 GTGNAACAGTGTGTCTGCCCAACCCAGGATGATGCTGAGCAGCAGACAGCTCTCTGG 1140

QY 381 IleserGlyTrpGlyAlaThrGluGluGlyGlyLysThrSerGluValLeuAenAlaAla 400
DB 1141 ATTTCCGGTGGGGGGCCACCGAGGAGAAAGGAGAACCTCAGAAAGTGTGAAACGCTGCC 1200

QY 401 LysValLeuLeuLeuGluThrGlnArqCysAenSerArgTyrrValTyrrAspAenLeuIle 420
DB 1201 AAGGTGCTTCTCATTTGACACAGAGATGCAACAGCAGATATGCTATGACAACTTGATC 1260

QY 421 ThrProAlaMetLysCysAlaGlyPheLeuGlnGlnAenValAspSerCysGlnGlyAsp 440
DB 1261 ACACGACCATGATCTGTGCGGCTTCTCTGAGGGGAAACGTCGATTTCTGCCAGGCTGAC 1320

QY 441 SerGlyProLeuValThrSerLysAenAenIleTtrProLeuIleGlyAspThrSer 460
DB 1321 AGTGGGGGCTCTGGTCACTTCGAGAAACAAATCTGGTGGCTGATAGGGGATACAAGC 1380

QY 461 TrpGlySerGlyCysAlaLysAlaTyrrArgProGlyValTyrrGlyAenValMetValPhe 480
DB 1381 TGGGGTCTGGCTGTGCAAGCTTACAGACGAGGAGTGTACGGAATGTGATGTATTTC 1440

QY 481 ThrAspTrpIleTyrrArgGlnMetArgAlaAspGly 492
DB 1441 ACGGACTGGATTTATCGACAATAGGGGCGACAGCGC 1476

```

RESULT 10

ABL95551 standard; cDNA; 1476 BP.

AC ABL95551;

DT 19-JUL-2002 (first entry)

DE Human P1000C cDNA sequence SEQ ID NO 931.

XX Human; cancer; prostate cancer; vaccine; cytostatic; immunostimulant;

KW gene therapy; gene; ss.

XX Homo sapiens.

OS US2002022248-A1.

PN

XX

PD 21-FEB-2002.

XX 12-JAN-2001; 2001US-00759143.

XX 25-FEB-1997; 97US-00806099.

PR 01-AUG-1997; 97US-00904804.

PR 10-FEB-1998; 98US-00020956.

PR 25-FEB-1998; 98US-00030607.

PR 14-JUL-1998; 98US-00115453.

PR 23-SEP-1998; 98US-00159812.

PR 15-JAN-1999; 99US-00232149.

PR 09-APR-1999; 99US-00288946.

PR 13-JUL-1999; 99US-00352616.

PR 12-NOV-1999; 99US-00439313.

PR 18-NOV-1999; 99US-00443686.

PR 14-JAN-2000; 2000US-00483672.

PR 27-MAR-2000; 2000US-00536857.

PR 09-MAY-2000; 2000US-00568100.

PR 12-MAY-2000; 2000US-00570737.

PR 13-JUN-2000; 2000US-00593793.

PR 27-JUN-2000; 2000US-00605783.

PR 10-AUG-2000; 2000US-00636215.

PR 29-AUG-2000; 2000US-00651236.

PR 06-SEP-2000; 2000US-00657279.

PR 02-OCT-2000; 2000US-00679426.

PR 10-OCT-2000; 2000US-00685166.

XX (XUJJ/) XU J.

PA (DILL/) DILLON D C.

PA (MITC/) MITCHAM J L.

PA (HARL/) HARLOCKER S L.

PA (JIAN/) JIANG Y.

PA (KALO/) KALOS M D.

PA (FANG/) FANGER G R.

PA (RETT/) RETTER M W.

PA (STOL/) STOLK J A.

PA (DAYC/) DAY C H.

PA (VEDV/) VEDVICK T S.

PA (CAET/) CARTER D.

PA (LISX/) LI S X.

PA (WANG/) WANG A.

PA (SKEI/) SKEIKY Y A W.

PA (HEPL/) HEPLER W T.

PA (HEND/) HENDERSON R A.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;

PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS,

PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;

XX WPI; 2002-255649/30.

XX New prostate-specific polynucleotides for diagnosing and treating

PT diseases, in particular prostate cancer, and as markers for the

PT progression of cancer.

XX Claim 1; SEQ ID NO 931; 87pp; English.

XX The present invention provides prostate-specific coding sequences and

CC their encoded proteins. These can be used in the diagnosis and treatment

CC of cancers, particularly prostate cancer. The present sequence is a cDNA

CC described in the invention

XX SQ Sequence 1476 BP; 361 A; 405 C; 402 G; 308 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3.21e-205 Length: 1476

Score: 2709.00 Matches: 491

Percent Similarity: 99.80% Conservative: 0

Best Local Similarity: 99.80% Mismatches: 1

Query Match: 99.71% Indels: 0

DB: Gaps: 0

US-09-615-285B-2 (1-492) x ABL95551 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Db 1 ATGGCTTTTGAACCTCAGGGTCACACCAGCTATTGGACCTTACTATGAAAACCATGATAC 60
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 Db 61 CAACCGGAACACCCCTATCCCGACACAGCCACTGTGGTCCCACTGTCTACAGATGCAT 120
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 121 CCGGCTCAGTACTACCCGTCCTCCCGTCGCCAGTAGCCGCCCGAGGGTCTGACGAGGCT 180
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 181 TCCAAACCCCGTGTCTGSCACGACGCCCAATCCCATCCGAGACAGTGTGCACCTCAAAG 240
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 Db 241 ACTAGAAAGCACTGTCATCACTTGACCTGGGACCTTCTCTGTGGAGCTCGGCTG 300
 QY 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 301 GCCGCTGGGCTACTCTCGAAGTTCATGGGCAGCAAGTCTCCAACTCTGGGATAGAGTGC 360
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 Db 361 GACTCCTCAGTACCTGCACTCAACCCCTCTAACTGTGTGATGGGTGTACACTGCCCC 420
 QY 141 GlyGlyLeuAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 421 GCGGGGAGGACAGAAATCGGTGTGTGCTTACGGATCAAACTTCATCCTTCAGGTG 480
 QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 Db 481 TACTCATCTCAGAGGAAGTCTGGCACCTGTGTGCCAAGACGACTGGAACGAGAACATAC 540
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 Db 541 GGGCGGGCGCTGTCAGGACATGGCTATAGAAATATTTTACTTAGCCACAGGAATA 600
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 601 GTGGATGACAGCGGATCCACCAAGCTTTATGAAACTCAACACAAAGTCCCGCAATTCGAT 660
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Db 661 ATCTATAAAACCTGTACCACAGTGTATGCTGTCTTCAAAAGCAGTGTGTCTTTACGC 720
 QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 Db 721 TGTATAGCTCGGGGTCAACTTGAACTCAAGCGCCAGCAGCAGGATTGTGGCGCGCAG 780
 QY 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
 Db 781 ACGCGCTCCCGGGGCGCTGGCCCTGGCGGTGAGCTGCAGCTCCAGAACGTCACGTG 840
 QY 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
 Db 841 TCGGAGGCTCCATCATCACCCCGAGTGTATCGTGACAGCGCCCACTCGGTGAAAAA 900
 QY 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 Db 901 CCTCTTTAAACATCCATGGCATTTGGACGGATTTGGCGGGATTTTGAGACAATCTTTTCATG 960
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 Db 961 TTCTATGGAGCCGATACCAAGTAGAAAAGTGAATTTCTCATCCCAATATGACTCCAAAG 1020
 QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 Db 1021 ACCAAGAACAAATGACATTCGCTGTATGAAGCTGCAGAGCCCTGCTCACTTTCAACGACCTA 1080

QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
 Db 1081 GTGAACACAGTGTGTCTGCCCAACCCAGGATGATGTGAGCCAGAACAGCTCTGCTGG 1140
 QY 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 Db 1141 ATTTCCCGGGTGGGGGCCACCGAGGAGAAAGGAACCTCAGAGTGTCTGACGCTGCC 1200
 QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 Db 1201 AAGTGTCTTCTATTGAGACACAGAGATGCAACAGCAGATATGTCTATGACCAACTGATC 1260
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 Db 1361 ACACCAAGCCATGATCTGTCCGGCTTCTGAGGGAACTCGATTTCTTCCAGGGTGAC 1320
 QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpIleLeuIleGlyAspThrSer 460
 Db 1321 AGTGAGGGCTCTGTGTCACCTCGAAGAACAAATATCTGTGTGGCTGATAGGGGATACAAGC 1380
 QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 Db 1381 TGGGTTTCTGGCTGTGCCAAAGCTTACAGACCAAGAGTGTACGGGAATGTGATGGTATTC 1440
 QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
 Db 1441 ACGGACTGGATTATCGACAAATGAGGGCAGACGGC 1476
 RESULT 11
 ACC95715
 ID ACC95715 standard; cDNA; 1476 BP.
 XX
 AC ACC95715;
 DT 28-AUG-2003 (first entry)
 XX
 DE Prostate tumour specific cDNA sequence SEQ ID 931.
 KW Cytostatic; gene therapy; prostate-specific protein; PSP; human;
 KW immune response; prostate cancer; ss.
 XX Homo sapiens.
 OS
 FN WO200289747-A2.
 XX
 PD 14-NOV-2002.
 XX
 PF 09-MAY-2002; 2002WO-US014753.
 XX
 PR 09-MAY-2001; 2001US-00852911.
 PR 29-JUN-2001; 2001US-00895814.
 PR 10-DEC-2001; 2001US-00012896.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Henderson RA;
 PI Kalos MD, Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS;
 PI Carter D, Li SX, Wang A, Skeiky YAW, Hepler WT, Hural J;
 PI McNeill PD, Houghton RL, Vinals Y De BassolisC, Foy TW, Watanabe Y;
 PI Deng T;
 XX
 DR WPI; 2003-167130/16.
 XX
 PT New prostate-specific proteins and genes, useful in gene therapy,
 PT particularly for stimulating an immune response in a patient, or treating
 PT prostate cancer in a patient, as well as for diagnosing prostate cancer
 PT in a patient.
 XX
 PS Example 2; Page 642; 691pp; English.
 XX
 CC The present invention relates to novel prostate-specific proteins (PSP)
 CC and their coding sequences. The PSPs and their coding sequences are
 CC useful for stimulating an immune response in a patient, or for treating

CC prostate cancer in a patient and for determining, detecting or diagnosing
CC the presence of a cancer in a patient. The present sequence was used to
CC illustrate the invention

XX SQ Sequence 1476 BP; 361 A; 405 C; 402 G; 308 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,21e-205 Length: 1476
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
Dbs: 7 Gaps: 0

US-09-615-285B-2 (1-492) x ACC95715 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 1 ATGGCTTTGAATCAGGCTCACCACAGCTATTGGACCTTACTATGAAACCATGGATC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
DB 61 CAACCGGAAACCCCTATCCCGCAGCCCACTGTGTGCCCATGTCTACGAGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CCGGCTCAGTACTACCCGTCGCCCGTGCACGAGTACGCCCGAGGGTCTTGACGAGGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 181 TCCAAACCCCGTGTCTGCAGCGAGCCCAATCCCATCCCGGAGCAGTGTGCACCTCAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB 241 ACTAAGAAGCACTGTGCATCACCTTGACCTGGGACCTTCTCGTGGAGCTGCCTG 300
QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB 301 GCCGCTGGCCTACTCTGGAAGTTCAATGGGCGACAGTGTCTCCAACTCTGGGATAGAGTGC 360
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
DB 361 GACTCCTCAGGTACTGTGCATCAACCCCTCTACTGGTGTGTGATGGGTGTGCATCTGCC 420
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
DB 421 GCGGGGAGGACGAGATCGGTGTGTGCTCTACCTCTACCGATCAAACTTCATCTTCAGGTG 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
DB 481 TACTCATCTCAGAGAAAGTCTGGCACCCTGTGTGCCAAGACGACTGGAACGAGAATAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
DB 541 GCGGGGCGGCTCTCAGGACATGGCTATAGAATAATTTTACTCTAGCCAAAGGAATA 600
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB 601 GTGGATGACAGCGGATCCACCAAGCTTTATGAACATGAACACCAAGTGGCGGAATGTGCAT 660
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 ATCTATAAATACTGTACCAAGTGTATGCTCTTCTTCAAAAGCAGTGTCTTTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
DB 721 TGTATAGCTGGGGTCAACTTGAACCTCAAGCCGACAGCAGGATTTGTGGGGCGGAG 780
QY 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
DB 781 AGCGGCTCTCCGGGGCCCTGGCCCTGGCAGTGCACGCTGCAGTGCAGACGCTCCACGTG 840
QY 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300

DB 841 TCGGAGGCTCCATCATCATCCCCCGAGTGGATCTGTGACAGCGCCCACTCGGTGAAAAA 900
QY 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
DB 901 CCTCTTACAATCCATGCGCATTTGGACGCAATTTGGGGGATTTGAGACAAATCTTCATG 960
QY 321 PheTyrGlyValaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
DB 961 TTCTATGAGCCGGATACCAAGTAGAAAAAGTGAATTTCTCATCAAAATTTATGACTCAAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
DB 1021 ACCAAGAACATGACATTTGCGCTGATGAAGCTGCAAGAGCCCTCTGACTTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
DB 1081 GTGAAACCAAGTGTGTCTGCCCAACCCAGGCATGATGTGTCAGCCAGACAGCTCTGCTGG 1140
QY 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
DB 1141 ATTTCGGGTGGGGGCCACCGAGGAGAAAGGAAACCTCAGAAGTGTGTAACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
DB 1201 AAGGTCTTCTCATTTGAGACAGAGATGCAACAGCAGATATGTCTATGACACCTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
DB 1261 ACACAGACCATGATCTGTGCGGCTTCTGTGAGGGGAAACGTCGATTTCTGCCAGGGTGC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
DB 1321 AGTGGAGGGCCTCTGTGCTCACTTCGAAGAAACAATATCTGGTGGCTGATAGGGATACAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
DB 1381 TGGGGTCTTGGCTGTGTCGCAAGCTTACAGACCCAGGAGTGTACGGGAATGTGATGGTATTC 1440
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
DB 1441 ACGGACTGGATTTATCGACAAATGAGGGCAGACGGC 1476
RESULT 12
ADBL4381
ID ADBL4381 standard; cDNA; 1476 BP.
XX
XX ADBL4381;
AC
XX 18-DEC-2003 (first entry)
DT
XX Human prostate specific cDNA P100C ORF (minus stop codon).
DE
XX Human; ss; prostate specific cDNA; cytostatic; immunostimulant;
KW gene therapy; cell therapy; vaccine; T-cell epitope;
KW class I major histocompatibility complex allele; MHC; prostate cancer;
KW tumour; antigen presenting cell.
XX
XX Homo sapiens.
XX
XX US2003185830-A1.
XX
XX 02-OCT-2003.
PD
XX 12-NOV-2002; 2002US-00294025.
PF
XX 25-FEB-1997; 97US-00806099.
XX 01-AUG-1997; 97US-00904804.
PR 01-AUG-1997; 98US-00020956.
PR 05-FEB-1998; 98US-00030607.
PR 14-JUL-1998; 98US-00115453.
PR 23-SEP-1998; 98US-00159812.
PR 15-JAN-1999; 99US-00232149.

PR 09-APR-1999; 99US-00288946.
 PR 13-JUL-1999; 99US-00352616.
 PR 12-NOV-1999; 99US-00439313.
 PR 18-NOV-1999; 99US-00443696.
 PR 14-JAN-2000; 2000US-00483672.
 PR 27-MAR-2000; 2000US-00536857.
 PR 09-MAY-2000; 2000US-00568100.
 PR 12-MAY-2000; 2000US-00570737.
 PR 13-JUN-2000; 2000US-00593793.
 PR 27-JUN-2000; 2000US-00605783.
 PR 09-AUG-2000; 2000US-00636215.
 PR 26-SEP-2000; 2000US-00851236.
 PR 02-OCT-2000; 2000US-00857279.
 PR 10-OCT-2000; 2000US-00679426.
 PR 09-NOV-2000; 2000US-00685166.
 PR 09-NOV-2000; 2000US-00709729.
 PR 12-JAN-2001; 2001US-00759143.
 PR 09-FEB-2001; 2001US-00780669.
 PR 09-MAY-2001; 2001US-00852311.
 PR 29-JUN-2001; 2001US-00895814.
 PR 10-DEC-2001; 2001US-00012896.
 PR 09-MAY-2002; 2002US-00144678.
 PR XX
 PA (CORI-) CORIXA CORP.

XX Xu J, Stolk JA, Kalos MD;

PI WPI; 2003-756193/71.

XX P-PSDB; ADB14382.

PT New isolated polypeptide for use in a vaccine for stimulating an immune
 PT response, or for treating or diagnosis cancer, preferably prostate
 PT cancer.

XX Example 2; Page; 101pp; English.

XX The invention relates to an isolated polypeptide comprising no more than
 CC 11-542 amino acids of ADB13563 comprising a sequence ADB14487. The
 CC peptides comprise a fragment ADB13563 of that contain naturally processed
 CC T-cell epitopes for 3 class I major histocompatibility complex (MHC)
 CC alleles. ADB13563 is a polypeptide encoded by a human prostate specific
 CC cDNA, one of 648 disclosed as new. Also included are nucleic acids
 CC encoding the proteins and peptides, expression vectors, a host cell
 CC transformed with the vector, an isolated antibody for antigen binding
 CC fragment that specifically binds to the protein or peptide, detecting
 CC the presence of a cancer in a patient (comprising contacting a patient
 CC sample with a binding agent that binds to the peptides or a polypeptide
 CC appearing as ADB13563, detecting the amount of polypeptide that binds to
 CC the agent and comparing the amount of polypeptide to a predetermined cut-
 CC off value to determine the presence of cancer), a fusion protein
 CC comprising the peptides or proteins, stimulating or expanding T cells
 CC specific for a tumor protein comprising contacting T cells with the
 CC peptides or the isolated T cell population, treating prostate cancer in a
 CC patient comprising administering a composition comprising the peptides,
 CC nucleic acids, antibodies or compounds, determining the presence of a
 CC cancer in a patient and treating prostate cancer in a patient comprising
 CC incubating cluster of differentiation (CD4+) and/or CD8+ T cells isolated
 CC from a patient with the peptides or antigen presenting cells that express
 CC the peptides so that the T cells proliferate, and administering the
 CC proliferated T cells to the patient. The peptides (or an oligonucleotide
 CC that hybridizes to nucleic acid encoding them), is used to detect the
 CC presence of cancer in a patient. The peptides, nucleic acids encoding, or
 CC antigen-presenting cells expressing the nucleic acid, are used to
 CC stimulate or expand T cells specific for a tumour protein. The peptides,
 CC nucleic acids, antibodies, fusion proteins, T cell populations or antigen
 CC presenting cells are used to stimulate an immune response or treat
 CC prostate cancer in a patient. The present sequence is a known cDNA
 CC showing sequence similarity to one of the disclosed human prostate
 CC specific cDNAs. Note: Except where otherwise indicated, the sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=20030185830.

XX

SQ Sequence 1476 BP; 361 A; 405 C; 402 G; 308 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3,21e-205 Length: 1476
 Score: 2709.00 Matches: 491
 Percent Similarity: 99.80% Conservatives: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 99.71% Indels: 0
 DB: 9 Gaps: 0

US-09-615-285B-2 (1-492) x ADB14381 (1-1476)

QY 1 MetAlaLeuAAsnSerGlySerProAlaIleGlyProTyrTyrGluAAsnHisGlyTyr 20
 Db 1 ATGGCTTTGAACCTCAGGTCACCCAGGCTATTGGACCTTACTATGAAACCACTGATAC 60
 QY 21 GlnProGluAAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 Db 61 CAACCGGAAACCCCTATCCGCACAGCCCACTGTGTCCTCCCACTGCTTACGAGGTGCAT 120
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 121 CCGGCTCAGTACTACCGCTCCCGTCCCGAGTACGCCCGGAGGCTCTGACGCGGCT 180
 QY 61 SerAAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 181 TCCAAACCCCGTCTCTGCACGCGAGCCCAATCCCATCCGGGACAGTGTGCACCTCAAG 240
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 Db 241 ACTAAGAAGCACTGTGCATCACTTGACCTGGGACCTTCTCTGTGGAGCTGGCTG 300
 QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAAsnSerGlyIleGluCys 120
 Db 301 GCGCTGGGCTACTCTGGAAGTTCATGGGCAAGTGTCTCAACTCTGGGATAGAGTGC 360
 QY 121 AspSerSerGlyThrCysIleAAsnProSerAAsnTyrCysAspGlyValSerHisCysPro 140
 Db 361 GACTCTCAGGTACTGTGATCAACCCCTCTAACTGTGTGTGTGTGTGTGTGTGTGTGTGT 420
 QY 141 GlyGlyGluAspGluAAsnArgCysValArgLeuTyrGlyProAAsnPheIleLeuGlnVal 160
 Db 421 GCGCGGGAGGAGCGAGAATCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
 QY 161 TyrSerSerGlnArgLysSerTyrPheProValCysGlnAAspTyrPheAAsnGluAAsnTyr 180
 Db 481 TACTATCTCAGAGGAGTCTTGGCACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAAsnPheTyrSerSerGlnGlyTle 200
 Db 541 GCGCGCGCGCGCTGTGAGGACATGGGCTATAGAAATAATTTTACTGTAGCGAAGAAATA 600
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAAsnThrSerAlaGlyAAsnValAsp 220
 Db 601 GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACCAAGTCCCGCAATGTGAT 660
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Db 661 ATCTATAAATAACTGTACCAAGT 720
 QY 241 CysIleAlaCysGlyValAAsnLeuAAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 Db 721 TGTATAGCTGTGGGGTCTACTGTGACTCAAGCCCGCAGACAGGATGTGTGGCGCGAG 780
 QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAAsnValHisVal 280
 Db 781 AGCGGCTCTCCGGGGGCTGTGGCGAGTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCT 840
 QY 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
 Db 841 TCGGAGGCTCCATCATCAACCCCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900
 QY 301 ProLeuAAsnProTyrPheIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320

Db 901 CCTCTTAACAAATCCATGGCATTTGACGGCATTTGCGGGATTTTGAGCAATCTTTCATG 960
Qy 321 PheTyrGlyAlaGlyTyrGlnValGluLeuValIleSerHisProAsnTyrAspSerLys 340
Db 961 TTCTATGAGCGCGATACCAAGTAGAAAAGTAGTTCTCATCCAAATATGACTCCAAG 1020
Qy 341 ThrLysAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACAAATGACATTTGGCGTGTATGAAGCTGCAGAACCTTGACTTTCAACGACCTA 1080
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
Db 1081 GTGAACACAGTGTCTGCCCAACCCAGGATGATGCTGCAGCCAGACAGCTCTGCTGG 1140
Qy 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCCGGTGGGGGCGCACGAGGAGAAAGGAGAACCTTCAGAACTGTGTGAACGCTGCC 1200
Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1201 ARGGTGCTTCTCATTTGAGACAGAGATGCAACAGCAGATATGCTATGACACCTGATC 1260
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACCAAGCCATGATCTGTGCGGCTTCTGCAGGGGAACGTCGATTCTTCCAGGGTGAC 1320
Qy 441 SerGlyClyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
Db 1321 AGTGGAGGGCTCTGTGTCCTTGAAGAACAAATCTGGTGGCTGATAGGGGATACAAGC 1380
Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGTTCCTGGCTGTGCAAGCTTACAGACCAAGGAGTGTACGGGAATGTGATGTTATTC 1440
Qy 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1441 ACGGACTGGATTATTCAGCAATGAGGGCAGACGGC 1476
RESULT 13
AAH93943
ID AAH93943 standard; cDNA; 1479 BP.
XX AC AAH93943;
XX DT 04-OCT-2001 (first entry)
XX DE P1000C open reading frame cDNA sequence with stop codon.
XX KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
XX KW cytosolic; gene therapy; metastasis; ss.
XX OS Homo sapiens.
XX PN WO200151633-A2.
XX PD 19-JUL-2001.
XX PF 16-JAN-2001; 2001WO-US001574.
XX PR 14-JAN-2000; 2000US-00483672.
XX PA (CORI-) CORIXA CORP.
XX PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
XX PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
XX PI Wang A, Meagher MJ;
XX DR WPI; 2001-425873/45.
XX New polynucleotide encoding a prostate-specific protein, for diagnosing,
PT monitoring and treating prostate cancer in a patient and for use in
PT vaccines.

XX Claim 1; Page 538; 543pp; English.
XX The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytosolic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated T cells
CC prepared using (I) or (II) are used to treat cancer in a patient. (I) and
CC the antibodies are also used in the detection of cancer in a patient. The
CC cancer that is diagnosed or treated is particularly prostate cancer. (I)
CC and (II) can be used in vaccines. The antibodies or (I) can be used for
CC monitoring the progression of cancer in a patient. (I) and (II) can also
CC be used to improve diagnostic and therapeutic methods for prostate
CC cancer. They can indicate the level of metastasis as well as the prostate
CC volume. AAH93357 to AAH93944 and AAH01115 to AAH01318 represent
CC polynucleotide and amino acid sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 1479 BP; 363 A; 405 C; 402 G; 309 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3, 21e-205 Length: 1479
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservations: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 4 Gaps: 0

US-09-615-285B-2 (1-492) x AAH93943 (1-1479)
Qy 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 1 ATGGTCTTGAATCAGGGTCCACCAGCTATTGGACCTTACTATGAAAACCATGGATAC 60
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
Db 61 CAACCGGAAACCCCTATCCCGCACGCCCACTGTGGTCCCCTGTCTACGAGTGCAT 120
Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACTACCCGTCGCCCGTCCCGCAGTACGCCCGGAGGTCTCAGCAGGCT 180
Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCCGTCTGTGACGAGCCCAATCCCATCCGGGACAGTGTGACCTCAAG 240
Qy 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 241 ACTAAGAAAGCACGTGTGCATCACCTTGACCTGGGACCTTCTCTGCGGAGCTGCGCTG 300
Qy 101 AlaAlaGlyLeuLeuTrpLysPheMetClySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCCGCTGGCCTACTCTGGAAGTTCTATGGGACAGAGTCTCCAACTCTGGGATAGAGTGC 360
Qy 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
Db 361 GACTCTCTCAGGTACCTGTGCATCAACCCCTCTAACTGGTGTGATGGGTGTCACTGCCCC 420
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleIleGlnVal 160
Db 421 GCGCGGAGGACGAGAATCGGTGTCTCCCTCTACCGATCAAACTTCTATCTTCAGGTG 480
Qy 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
Db 481 TACTCATCTCAGAGGAAGTCTTGGACCTCTGTGTCGAAGACGACTGGAACGAGAATAC 540
Qy 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 541 GCGCGGCGGCTCGCAGGACATGGGCTATAAGATATAATTTTACTCTAGCCCAAGATA 600
Qy 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 601 GTGGATGACGCGGATCCACGCTTTTATGAACCTTTATGAACACACAGTGGCGGCAATCTGCAT 660

```
QY 221 IletYrLysLysLeuYrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATAAAACCTGACACAGTGATCGCTGTTCTTCAAAGCAGTGCTTCTTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 721 TGTATAGCCTCGGGGCTCACTTGAACTCAAGCCGACAGACAGATGTGGGCGGAG 780
QY 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 781 AGCGCGCTCCCGGGGCTGGCCCTGGCAGTGCAGCTGCAGCTGCAGAACGTCCACGTG 840
QY 281 CysGlyGlySerIleIleThrProGlnTrpIleValThrAlaAlaHisCysValGlyLys 300
Db 841 TCGGAGGCTCCATCATCACCCCGAGTGGATCTGCAGCGCCGCTGCTGGTGGAAAA 900
QY 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTACATCATCGATGCGATTTGGAGCGATTTGGGGGATTTGAGACATCTTTTCATG 960
QY 321 PheYrGlyAlaGlyYrGlnValGlyLysValIleSerHisProAsnTyAspSerLys 340
Db 961 TTCTATGAGCGGATACCAAGTAGAAAAAGTGAATTTCTCATCCAAATTTATGACTCCAAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACATGACATTCGCGTGTATGAGCTGCAGAAAGCTCTGACTTTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1081 GTGAACACAGTGTGTCGCCAACCCAGCGCATGCTGCAGCCAGAACACCTCTGCTGG 1140
QY 381 IleSerGlyTrpGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCCGGTGGGGGCGCCACCGAGAGAAAGGAGACCTCAGAAAGTCTGAACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTrpValTyAspAsnLeuIle 420
Db 1201 AAGTGTCTTCTATGACACAGAGATGCACAGCATATGTCATATGACAACTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACACGACCATGATCTGTGCGGCTTCTGCGAGGGAACGTGATTTCTGCGAGGCTGAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1321 AGTGAAGGCGCTCTGCTGCTACCTTCGAAGAACAAATATCTGCTGCTGATAGGGGATACAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyArgProGlyValTyArgGlyAsnValMetValPhe 480
Db 1381 TGGGGTCTGCTGTGCCAAAGCTTACAGACCAGGAGTGTACGGGAATGTGATGTTATTC 1440
QY 481 ThrAspTrpIleTyArgGlnMetArgAlaAspGly 492
Db 1441 ACGGACTGATTTATCGACAAATGAGGCGACAGCGC 1476
RESULT 14
AAS64179
ID AAS64179 standard; cDNA; 1479 BP.
XX AC AAS64179;
XX DT 29-JAN-2002 (first entry)
XX DE Human prostate cDNA sequence #598.
XX KW Human; prostate cancer; ss; cytostatic; immunostimulant; tumour.
XX OS Homo sapiens.
XX PN WC200173032-A2.
XX
```

```
PD 04-OCT-2001.
XX 27-MAR-2001; 2001WO-US009919.
XX 27-MAR-2000; 2000US-00536857.
PR 03-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
PR 27-JUN-2000; 2000US-00605783.
PR 09-AUG-2000; 2000US-00636215.
PR 26-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
PR 03-NOV-2000; 2000US-00709729.
XX (CORI-) CORIXA CORP.
PA
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GS, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX WP1; 2001-639232/73.
DR P-PSDB; AAU69960.
DR
XX New human prostate-specific polypeptides and polynucleotides useful for
PT the diagnosis and treatment of cancer, especially prostate cancer.
PT
XX Claim 1; Page 571-572; 579pp; English.
PS
XX The invention relates to isolated prostate-specific polynucleotides,
CC polypeptides, fusion proteins of the polypeptides, antibodies raised
CC against the polypeptides (or antigenic epitopes derived from them) and
CC antigen-presenting cells expressing the polypeptides. The antibodies are
CC useful for detecting the presence of cancer, especially prostate cancer.
CC The polypeptides, polynucleotides and the antigen-presenting cells are
CC useful for stimulating and/or expanding T cells specific for a tumour
CC protein, and for inhibiting the development of cancer especially prostate
CC cancer. Compositions comprising the polynucleotide and/or polypeptide are
CC useful for stimulating an immune response, and for treating cancer. The
CC oligonucleotide is useful for detecting cancer. The present sequence is a
CC prostate specific polynucleotide of the invention
XX
SQ Sequence 1479 BP; 363 A; 405 C; 402 G; 309 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 3-21e-205 Length: 1479
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 4 Gaps: 0
US-09-615-285B-2 (1-492) x AAS64179 (1-1479)
QY 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyTrpGluAsnHisGlyTyr 20
Db 1 ATGGCTTTGAACCTCAGGGTACCACAGCTATTGGACCTTACTATGAAACCATGATAC 60
QY 21 GlnProGluAsnProTyTrpProAlaGlnProThrValProThrValTyTrpGluValHis 40
Db 61 CAACCGGAAACCCCTATCCGACAGCCACTGTGTCCTCCCACTGCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyTrpProSerProValProGlnTyTrpAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACTACCCGTCCTCCCGCCAGTACGCGCCGAGGGTCTTCGACGAGGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCGCTGCTGTCAGCAGAGCCCAATCCCATCCGGGACAGTGTGCACCTCAAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
```


Db 241 ACTAAGAAAGCACTGTGCATCATCTTGACCTGGGACCTTCTCTGTGGAGCTGGCTG 300
Qy 101 AlaalaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCGCTGGCTACTCTCGAAGTTTCATGGGAGCAAGTGTCTCCAACTCTGGATAGAGTGC 360
Qy 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
Db 361 GACTCTCAGGTACTGTGCATCAACCTCTTAACCTGGTGTGATGGCTGTGCACACTGCC 420
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTrpGlyProAsnPhelleLeuGlnVal 160
Db 421 GCGGGGAGGACGAGAAATCGGTGTGTCTGCGCTCTACGGATCAAACTTCATCTTCAGGTG 480
Qy 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTrp 180
Db 481 TACTCTCTCAGAGAGTCTTGGCACCTGTGTGCCAAGCAGCTGGACGAGAACTAC 540
Qy 181 GlyArgAlaAlaCysAAspMetGlyTyrIlyAsnAsnPheTyrSerSerGlnGlyIle 200
Db 541 GCGGGGCGGCTCGCAGGACATGGGTATAAGAAATAATTTTACTCTAGCCCAAGAAATA 600
Qy 201 ValAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 601 GTGATGACAGCGGATCCACAGCTTTATGAACCTGAACAAAGTGCAGGCAATGTGAT 660
Qy 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATAAAGAACTGTACACAGTATGCTGCTTCTTCAAAAGCAGTGGTTCCTTACGC 720
Qy 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 721 TGTATAGCTCGGGGGTCAACTTGAACCTCAAGCGCCAGAGCGAGATGTGGCGCGGAG 780
Qy 261 SerAlaLeuProGlyAlaTrpTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 781 AGCGGCTCCGGGGGCTGGCTTGGAGGTGAGCTGACCTGCAGTCCAGAACGTCACGCTG 840
Qy 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db 841 TCGGAGGCTCATCATCACCCCGAGTGTGATCGTGACGCGGCCACTCGCTGGAAAAA 900
Qy 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTTAACATTCATGGCATTTGACGGCATTTGCGGGGATTTTGACAACTCTTTCATG 960
Qy 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 961 TTCTATGAGCGGATACCAAGTAGAAAAAGTATTTCTCATCCAAATTTATGACTCCAA 1020
Qy 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACATGATTTGCCCTGATGAGCTGCAGAGCTCTGACTTTCAACGACTTA 1080
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1081 GTGAACACAGTGTCTGCCCAACCCAGGATGATGCTGCAGCCAGAACAGCTCTGTGTTG 1140
Qy 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCCGGGTGGGGGCCACCGAGAGAAAGGAGACCTCAGAAAGTGTGAACGCTGCC 1200
Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1201 AAGGTGCTTCTCATTTGACACACAGATGCAACAGCAGATATGCTATGCAACCTGATC 1260
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACGAGCATGATCTGTGGCGGCTTCTGTGAGGGGAAACGTCGATTTCTGCGAGGTGAC 1320
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1321 AGTGGAGGCGCTCTGGTCACTTCGAGAAACAATATCTGGTGGCTGATAGGGGATACAGC 1380

Qy 461 TrpGlySerGlyCysAlaLysAlaLysValTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGTTCTGGCTGTGCCAAAGCTTACAGACGAGAGTGTACGGGAATGTGATGTATTC 1440
Qy 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1441 ACGGACTGGATTATTCGACAAATGAGGCGACAGCGC 1476

RESULT 15

ACAS9987
ID ACAS9987 standard; cDNA; 1479 BP.

XX ACAS9987;

XX 10-JUN-2003 (first entry)

XX Prostate cancer therapy associated cDNA #668.

XX Prostate cancer; vaccine; gene therapy; cytostatic; fusion protein;
KW immunogen; cancer; prostate specific antigen; PSA;
KW prostatic acid phosphatase; PAP; prostate specific membrane antigen;
XX PSMA; gene; ss.

XX Homo sapiens.

XX US2002192763-A1.

XX 19-DEC-2002.

XX 29-JUN-2001; 2001US-00895793.

XX 04-OCT-1999; 99US-0157455P.

XX 04-OCT-2000; 2000US-00679272.

XX 28-MAR-2001; 2001US-00822827.

XX (XUJ/) XU J.

XX (DILL/) DILLON D C.

XX (MITC/) MITCHAM J L.

XX (HARK/) HARLOCKER S L.

XX (JIAN/) JIANG Y.

XX (KALO/) KALOS M D.

XX (FANG/) FANGER G R.

XX (RETT/) RETTER M W.

XX (STOL/) STOLK J A.

XX (DAYC/) DAY C H.

XX (VEDV/) VEDVICK T S.

XX (CART/) CARTER D.

XX (LISX/) LI S X.

XX (WANG/) WANG A.

XX (SKEI/) SKEIKY Y A W.

XX (HEPL/) HEPLER W T.

XX (HEND/) HENDERSON R A.

XX (HURA/) HURAL J.

XX (MCNE/) MCNEILL P D.

XX (HOUG/) HOUGHTON R L.

XX (DBAS/) Y DE BASSOLS C V.

XX (FOYT/) FOY T M.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Y De Bassols C, Hural J;
PI McNeill PD, Houghton RL, Y De Bassols C, Foy TM;
XX WPI; 2001-245062/25.

XX Prostate specific protein and its encoding polynucleotide, useful for the
PT treatment and diagnosis of prostate cancer.

XX Example 2; SEQ ID NO 930; 85pp; English.

XX The invention describes a fusion protein comprising at least one amino
CC acid sequence of immunogenic portions of any of the 3 sequences not

CC defined in the specification, or sequences having at least 70 or 90 %
CC sequence identity to any one of the 35 sequences defined in the USPTO web
CC site, which is encoded by any of the 4 nucleotide sequences not defined
CC in the specification. The fusion protein, composition and methods are
CC useful for diagnosing, preventing and/or treating cancer, particularly
CC prostate cancer. The proteins are useful as markers to indicate the
CC presence or absence of cancer. This sequence represents a prostate cancer
CC therapy associated cDNA. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=US20020192763
XX
SQ Sequence 1479 BP; 363 A; 405 C; 402 G; 309 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,21e-205 Length: 1479
Score: 2709.00 Matches: 451
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 5 Gaps: 0

US-09-615-285B-2 (1-492) x ACA59987 (1-1479)

QY	1	MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr	20
DB	1	ATGGCTTTGAACCTCAGGGTCAACACAGCTATTGGACCTTACTATGAAACCATGGATAC	60
QY	21	GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis	40
DB	61	CAACCGGAAACCCCTATCCGGCAGCCACCTGTGTGCTCCACCTGTCTACGAGGTGCAT	120
QY	41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60
DB	121	CGGGCTCAGTACTACCGTCCCGGTCGCCAGTACGCCCGCGAGGGTCTCGACGAGGCT	180
QY	61	SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys	80
DB	181	TCCAAACCCCGTGTCTGCGAGAGCCCAATCCCATCCGGGACAGTGTGCACTCAAG	240
QY	81	ThrLysLeuAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu	100
DB	241	ACTAAGAAAGCATGTGCATCACTTGACCTGGGGACCTTCTCTGTGGGAGCTGGCTG	300
QY	101	AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys	120
DB	301	CGCGCTGGCCCTACTCTGGAAGTTTCATGGGCAGCAAGTGTCCAACCTCTGGGATAGAGTGC	360
QY	121	AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro	140
DB	361	GACTCTCAGGTACCTGCATCAACCCCTCTAACCTGTGTGTGGGTGTGCACACTGCCCC	420
QY	141	GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal	160
DB	421	GGCGGGGAGGACGAGAAATCGGTGTGTGTGCTCTACGGATCAAACTTCACTCTCAGGTG	480
QY	161	TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspTrpAsnGluAsnTyr	180
DB	481	TACTCATCTCAGAGGAAGTCTGGCCCTGTGTGCCAAGACGACCTGGAAACGAGAACTAC	540
QY	181	GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle	200
DB	541	GGCGGGCGGCTCGAGGACATGGGCTATAGAAATATTTTACTCTAGCCAAAGGAATA	600
QY	201	ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp	220
DB	601	GTGGATGACAGCGGATCCACAGCTTTATGAAACTGAACACAGTCCCGGCAATGTCGAT	660
QY	221	IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg	240
DB	661	ATCTATTAATAAATGTGTACCACTGATGCTGTCTTCTCAAAAGCAGTGTCTTTTACGC	720
QY	241	CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu	260

Search completed: June 1, 2004, 14:59:22
Job time : 668 secs

DB	721	TGTATAGCTCGCGGGTCAACTTGAACCTCAAGCCGCGAGAGGATTTGTGGCGCGAG	780
QY	261	SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal	280
DB	781	AGCCGCTCCCGGGGCTGGCCCTGGCAGTCAAGCTGCACGTCAGAACGTCACGCTG	840
QY	281	CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys	300
DB	841	TGCGAGGCTCCATCATCAACCCCGAGTGGATCGTACAGCGCCCACTGCGTGAAAAA	900
QY	301	ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet	320
DB	901	CCTCTTAACAATCATGGCATTTGGACGGCATTTTGGGGGATTTTGAGACAACTTTTCATG	960
QY	321	PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys	340
DB	961	TTCTATGAGCGCGGATACCAAGTAGAAAAAGTAGTTTCTCATCCAATTTATGATCTCAAG	1020
QY	341	ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu	360
DB	1021	ACCAAGAACCAATGATGATGGCTGATGAAGCTGCAGAAAGCTCTGACTTTCAACGACCTA	1080
QY	361	ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp	380
DB	1081	GTGAACCAAGTGTCTGCCCAACCCAGGATGATGCTGCAGCCAGAACAGCTCTGCTGG	1140
QY	381	IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla	400
DB	1141	ATTTCCGGTGGGGGCCACCCAGAGAGAAAGGGAAGACCTCAGAAAGTGTCTGAACGCTGCC	1200
QY	401	LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle	420
DB	1201	AAGTGTCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGTCTATGACCACTGATC	1260
QY	421	ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp	440
DB	1261	ACACAGCCATGATCTGTGCGGCTTCTGCAGGGGAACGTCGATTCTTGCAGGGGTGAC	1320
QY	441	SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer	460
DB	1321	AGTGGAGGGCTCTGGTCACTTCGAAACAATATCTGGTGGCTGTAGGGGATACAAAGC	1380
QY	461	TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe	480
DB	1381	TGGGTTCTGGCTGTGCAAAAGCTTACAGACCAGAGTGTACGGGAATGTGATGTTATTC	1440
QY	481	ThrAspTrpIleTyrArgGlnMetArgAlaAspGly	492
DB	1441	ACGGACTGGATTTTATCGACAAATGAGGGCAGACGGC	1476

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 1, 2004; 14:39:10 ; Search time 5848 Seconds
(without alignments)
3646.507 Million cell updates/sec

Title: US-09-615-285B-2
Perfect score: 2717
Sequence: 1 MALNSGPPAIGPYENHGY.....VGNVNVFTDWIYQNRADG 492

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09615285/runat_01062004_132133_19946/app_query.fasta_1.647
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTWT=ptc -NORM=ext -HRAPISE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09615285 @CGN_1_1_3731 @runat_01062004_132133_19946 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rcd.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em.sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2717	100.0	1479	6	AR25345	Sequence
2	2717	100.0	1738	6	AX395316	Sequence
3	2717	100.0	1738	6	BD218220	Tumor ant
4	2717	100.0	1740	9	AF270487	BC051839 Homo sapi
5	2717	100.0	3195	9	BC051839	Homo sapi
6	2714	99.9	1479	9	AF123453	Homo sapi
7	2711	99.8	3226	9	AF329454	Homo sapi
8	2709	99.7	1476	6	AX201157	Sequence
9	2709	99.7	1476	6	AX201157	Sequence
10	2709	99.7	1479	6	AX201157	Sequence
11	2709	99.7	1479	6	AX201157	Sequence
12	2709	99.7	3245	6	AX201156	Sequence
13	2709	99.7	3245	6	AX267955	Sequence
14	2704	99.5	3244	6	BD244778	Specific
15	2704	99.5	3443	6	BD228306	Method of
16	2696	99.2	2479	6	BD235838	A novel m
17	2696	99.2	2479	6	AR225372	Sequence
18	2696	99.2	2479	6	AR405834	Sequence
19	2696	99.2	2479	6	AX201121	Sequence
20	2696	99.2	2479	6	AX207965	Sequence
21	2696	99.2	2479	6	AX267920	Sequence
22	2696	99.2	2479	6	AX306771	Sequence
23	2696	99.2	2479	6	AX395318	Sequence
24	2696	99.2	2479	6	BD218221	Tumor ant
25	2696	99.2	2479	9	HSU75329	Human serin
26	2696	99.2	3966	6	AX041973	Sequence
27	2121	78.1	3175	10	BC054348	Mus muscu
28	2119	78.0	3180	10	BC061712	Rattus no
29	2117	77.9	1735	10	AF243500	Mus muscu
30	2117	77.9	3014	10	BC038393	Mus muscu
31	2116	77.9	1753	10	AF113596	Mus muscu
32	2107	77.5	1755	10	AF199362	Mus muscu
33	2083	76.7	1473	10	AB073550	Rattus no
34	1569	57.7	1077	6	AR194808	Sequence
35	1484	54.6	1001	6	AR270400	Sequence
36	1165	42.9	683	6	AR405835	Sequence
37	1165	42.9	683	6	AX201123	Sequence
38	1165	42.9	683	6	AX267922	Sequence
39	1009	37.1	3507	5	AB038497	Xenopus l
40	894	32.9	2948	6	AX354825	Sequence
41	891.5	32.8	2412	6	AX375996	Sequence
42	891.5	32.8	2412	9	AY358458	Homo sapi
43	891.5	32.8	2418	6	BD173589	Novel ser
44	885	32.6	1359	6	AX354827	Sequence
45	885	32.6	1359	6	BD173588	Novel ser

ALIGNMENTS

RESULT 1

AR225345
LOCUS AR225345 1479 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 1 from patent US 644419.
ACCESSION AR225345
VERSION AR225345.1 GI:27263288
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1479)
AUTHORS Wong,A.K.C., Tavtigian,S.V. and Teng,D.H.F.
TITLE TWPRSS2 is a tumor suppressor
JOURNAL Patent: US 644419-A 1 03-SEP-2002;
FEATURES Location/Qualifiers
1..1479
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 3,14e-197 Length: 1479
Score: 2717.00 Matches: 492
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-615-285B-2 (1-492) x AR225345 (1-1479)

QY 1 MetAlaLeuAenSerGlySerProAlaIleGlyProTyrTyrGluAenHisGlyTyr 20
Db 1 ATGGCTTTGAACCTCAGGCTCAGGCTCAGGCTCAGGCTCAGGCTCAGGCTCAGGCT 60
QY 21 GlnProGluAenProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 61 CAACCGGAAACCCCTATCCCGCACGCCCACTGTGTCCCACTGTCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACTACCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCT 180
QY 61 SerAenProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCCGTGTGTGACGACGACCCCAATCCCAATCCCGGACAGTGTGCACCTCAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
Db 241 ACTAAGAAAGCAGCTGTGCATCACCTTGACCTCGGACCTTCCTCGTGGAGCTCGCGTG 300
QY 101 AlaAlaGlyLeuLeuTyrPhePheMetGlySerLysCysSerAenSerGlyIleGluCys 120
Db 301 GCCGCTGCCCTACTCTGGAAGTTTCATGGGACAGAGTCTCCAACTCTGGGATAGAGTGC 360
QY 121 AspSerSerGlyThrCysIleAenProSerAenTyrCysAspGlyValSerHisCysPro 140
Db 361 GACTCTCAGGTACCTGTGCATCAACCCCTTAACCTGTGTGTGTGTGTGTGTGTGTGT 420
QY 141 GlyGlyGluAspGluAenArgCysValArgLeuTyrGlyProAenPheIleLeuGlnVal 160
Db 421 GGCGGGGAGGACGAGATCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTyrAenGluAenTyr 180
Db 481 TACTCATCTCAGAGGAAGTCTCGGACCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAenAenPheTyrSerSerGlnGlyIle 200
Db 541 GGCGGGGCGGCTCGAGGACATGGGCTATAGAAATAATTTTACTCTAGCCCAAGGAATA 600
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAenThrSerAlaGlyAenValAsp 220
Db 601 GTGGATGACAGCGGATCCACAGCTTTATGAAACTGAAACAGTGCCTGGCAATGTGCAT 660

QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATATAAAACCTGTACCAAGTGTGCTGTCTTTCTTCAAAAGCAGTGTGTCTTACGC 720
QY 241 CysIleAlaCysGlyValAenLeuAenSerSerArgLysArgIleValGlyGlyGlu 260
Db 721 TGTATAGCTCGGGGCTCAACTTGAACCTCAAGCCCGCAGAGCAGATCTGTGGGCGGAG 780
QY 261 SerAlaLeuProGlyValAlaTyrProTyrGlnValSerLeuHisValGlnAenValHisVal 280
Db 781 AGCGGCTTCCCGGGGCTGTGCTGTGAGGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 840
QY 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
Db 841 TGGGAGGCTCCATCATCACCCCGAGTGGATCTGACAGCCGCCCTCTGACTTTCAACGACCTA 900
QY 301 ProLeuAenAenProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTTAACAATCCATGACATTTGGAGGCAATTTGGGGGATTTTGAGACATCTTTTCATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAenTyrAspSerLys 340
Db 961 TTCTATGAGCGCGGATACCAAGTAGAAAAGTGTATTTCTATCCAAATATATGACTCCAAAG 1020
QY 341 ThrLysAenAenAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAspLeu 360
Db 1021 ACCAAGAAACAATGACATTTGGCTGATGAAGCTGAGAAAGCTCTGACTTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAenProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
Db 1081 GTGAACACAGTGTGTCTGCCAACCCAGCATGATGTGTGACGCCAGAAACAGCTGTGTGG 1140
QY 381 IleSerGlyTyrGlyValThrGluLysGlyLysThrSerGluValLeuAenAlaAla 400
Db 1141 ATTTCCGGGTGGGGGCCACCCAGGAGAAAGGAAAGACTCAGAAAGTGTGTGACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAenSerArgTyrValTyrAspAenLeuIle 420
Db 1201 AAGTGTCTTCTCATTTGAGACACAGAGATGCAACAGACATATGTCTATGACAACTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAenValAspSerCysGlnGlyAsp 440
Db 1261 ACACCCAGCATGATCTGTGCCGCTTCTGTGACGGGAAAGCTGTGTGTGTGTGTGTGTGT 1320
QY 441 SerGlyGlyProLeuValThrSerLysAenAenIleTyrTyrPheIleGlyAspThrSer 460
Db 1321 AGTGAGGGCCCTGTGTCTACTTCGAAAGCAATATCTGTGTGTGTGTGTGTGTGTGTGT 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAenValMetValPhe 480
Db 1381 TGGGGTCTGT 1440
QY 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
Db 1441 ACGGACTGTGATTATCGAATAATGAGGCGAGCGGC 1476

RESULT 2
AX395316 1738 bp DNA linear PAT 18-MAY-2002
LOCUS AX395316
DEFINITION Sequence 1 from Patent WO0204953.
ACCESSION AX395316
VERSION AX395316.1 GI:21066313
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Safran,D., Raitano,A.B., Hubert,R.S., Jakobovits,A., Faris,M. and Challita-Eid,P.M.
TITLE Novel tumor antigen useful in diagnosis and therapy of bladder, ovary, lung and kidney cancers
JOURNAL Patent: WO 0204953-A 1 17-JAN-2002;

Agensys, Inc. (US)
 Location/Qualifiers
 1. 1738
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Alignment Scores:
 Pred. No.: 3.81e-197 Length: 1738
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-615-285B-2 (1-492) x AX395316 (1-1738)

QY 1 MetAlaLeuAenSerGlySerProAlaIleGlyProTyrTyrGluAenHisGlyTyr 20
 DB 112 ATGGCTTTGAACCTAGGGTCAACACAGCTATTGGACCTTACTATGAACACCTGGATAC 171
 QY 21 GlnProGluAenProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 DB 172 CAACCGAAACCCCTATCCGACACAGCCCACTGTGGTCCCACTGTCTACGAGGTGCAT 231
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 DB 232 CCGGCTCAGTACTACCGCTCCCGGTGCCAGTACGCCCGAGGGTCTGTACACAGCT 291
 QY 61 SerAenProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 DB 292 TCCAAACCCGCTGCTGCACGAGCCCAATCCCAATCCCGACAGTGTGCACCTCAAG 351
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 DB 352 ACTAAGAAAGACCTGTGCATCACTTACCTCGGGACCTTCTCGTGGAGCTGCCCTG 411
 QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAenSerGlyIleGluCys 120
 DB 412 CCGGCTGCCCTACTCTGAGTTCATGGGACAGTGTCTCAACTCTGGATAGATGC 471
 QY 121 AspSerSerGlyThrCysIleAenProSerAenTyrCysAspGlyValSerHisCysPro 140
 DB 472 GACTCCTCAGGTACCTGCATCAACCCCTCTAACTGGTGTGATGGCGTGCACATGCC 531
 QY 141 GlyGlyGluAenGluAenArgCysValArgLeuTyrGlyProAenPheIleLeuGlnVal 160
 DB 532 GCGGGGAGGACGAGAACTGGTGTTCCTCTACGACCACTTCACTTCTCAGGTG 591
 QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAenAspTrpAenGluAenTyr 180
 DB 592 TACTCATCTCAGAGGAAGTCTCTGGCACCTGTGTGCCAAGACGACTGGAACGAGAATAC 651
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAenAenPheTyrSerSerGlnGlyLe 200
 DB 652 GCGCGGGGGGCTGTGAGGACATGGCTATAGAAATAATTTTACTCTAGCCAAAGGAATA 711
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAenThrSerAlaGlyAenValAsp 220
 DB 712 GTGGATGACGCGGATCCACCGCTTTTGAACCTGAACACAGTGGCGGCAATGTGCAT 771
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 DB 772 ATCTATAAAAACTTACCACAGTATGCTGTCTTCAAAAAGCAGTGGTTCCTTACGC 831
 QY 241 CysIleAlaCysGlyValAenLeuAenSerSerArgGlnSerArgIleValGlyGlyGlu 260
 DB 832 TGATAGCTGCGGGGTCAACTTGAACCTCAAGCCCGCAGCAGGANTTGTGGCGGCGAG 891
 QY 261 SerAlaLeuProGlyValaTyrProTyrGlnValSerLeuHisValGlnAenValHisVal 280
 DB 892 ASCGCGCTCCCGGGGCTGGCCCTGGCAGGTGACCTGCAGTCCAGAACGTCCAGCTG 951

QY 281 CysGlyGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
 DB 952 TCGGAGGCTCCATCATCCCCCGAGTGGATCGTGACAGCGCCACCTGCGTGGAAAAA 1011
 QY 301 ProLeuAenAenProTyrPheIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 DB 1012 CTCTTTAACAATCCATGGCATTTGACGCGATTTGCGGGATTTTGAGACAACTTTTCATG 1071
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAenTyrAspSerLys 340
 DB 1072 TTCTATGGAGCGGATACCAAGTAGAAAAAGATTTCTCATCCAAATATGACTCCAAAG 1131
 QY 341 ThrLysAenAenAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAspLeu 360
 DB 1132 ACCAAGAACAATGACATTTGCGCTGATGAGCTGCAGAGCCCTGACTTTCAACGACCTA 1191
 QY 361 ValLysProValCysLeuProAenProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
 DB 1192 GTGAACACAGTGTCTGTGCCAACCCAGCATGATGCTGCAGCCAGACAGCTCTGCTGG 1251
 QY 381 IleSerGlyTyrGlyValaThrGluGluLysGlyLysThrSerGluValLeuAenAlaAla 400
 DB 1252 ATTTCGGGTGGGGGCCACCCAGGAGAAAGGAGACCTCAGAGTGTCTGAACGCTGCC 1311
 QY 401 LysValLeuLeuIleGluThrGlnArgCysAenSerArgTyrValTyrAspAenLeuIle 420
 DB 1312 AAGTGTCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGCTATGACAACTGATC 1371
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAenValAspSerCysGlnGlyAsp 440
 DB 1372 ACACCAAGCATGATCTGTCCGCTTCTTCAGAGGAGACGTGATTTCTTCCAGGGTGAC 1431
 QY 441 SerGlyGlyProLeuValThrSerLysAenAenIleTyrTrpLeuIleGlyAspThrSer 460
 DB 1432 AGTGGAGGCTCTGTGTCTTCCGAGAACATATCTGTGTGCTGATAGGGGATACAAGC 1491
 QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAenValMetValPhe 480
 DB 1492 TGGGGTCTCTGGCTGTGCCAAGCTTACAGACAGGAGTGTACGGGAATGTGATGTTTC 1551
 QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
 DB 1552 ACGGACTGTGATTTATGCAAAATGAGGGGACAGCGC 1587

RESULT 3
 BD218220 1738 bp DNA linear PAT 17-JUN-2003
 LOCUS
 DEFINITION
 Tumor antigen useful in diagnosis and therapy of prostate and colon cancer.

BD218220
 VERSION
 BD218220.1 GI:33027990
 KEYWORDS
 JP 2002517185-A/1.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 1738)
 AFAR,D.E., Hubert,R.S., Leong,K., Raitano,A.B., Saffran,D.C. and
 Mitchell,S.C.
 Tumor antigen useful in diagnosis and therapy of prostate and colon
 cancer
 Patent: JP 2002517185-A 1 18-JUN-2002;
 UROGENESIS INC
 OS Homo sapiens (human)
 PN JP 2002517185-A/1
 PD 18-JUN-2002
 PF 01-JUN-1999 JP 2000552152
 PR 01-JUN-1998 US 60/087598,29-JUN-1998 US 60/091474 PR
 14-APR-1999 US 60/129521
 PI DANIEL E AFAR, RENE S HUBERT, KAHAN LEONG, ARTHUR B RAITANO PI
 DOUGLAS C SAFFRAN,
 PI STEPHEN CHAPPELL MITCHELL

PC C12N9/64,A61K38/00,A61K39/395,A61K39/395,A61P13/08,A61P35/00,
 PC C07K14/47,
 PC C07K16/40,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12N15/09 PC
 C12P21/08,C12Q1/68,
 PC G01N33/50,G01N33/566,G01N33/574,A61K37/02,C12N5/00,C12N15/00
 CC Tumor antigen useful in diagnosis and therapy of prostate and
 CC colon cancer
 FH Key Location/Qualifiers
 FT CDS (112)...(1598).

FEATURES
 source
 1. .1738
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 3,81e-197 Length: 1738
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservatve: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-615-285B-2 (1-492) x BD218220 (1-1738)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Db 112 ATGGCTTTGAAGTCAAGGTCACACAGCTATTGGACCTTACTATGAAACCATGATAC 171
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 Db 172 CAACCGGAAACCCCTATCCCGACAGCCCACTGTGTGCCCTCTCTACGAGGTGCAT 231
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 232 CCGGCTCAGTACACCGCTCCCGTCCCGTCCAGTACCGCCCGAGGTCCTCGACCGAGGT 291
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 292 TCCAAACCCGTCGTCACGACGAGCCCAATCCCATCCGAGCAGTGTGCACCTCAAG 351
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 Db 352 ACTAAGAAAGCAGTGTGCATCACCTTGACCTGGGGACCTCTCTCGTGGAGTGGCTG 411
 QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 412 GCCGCTGGCCTACTCTGAAAGTTCATGGGAGCAAGTGTCTCCAACTCTGGGATAGAGTGC 471
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 Db 472 GACTCCTCAGGTACCTGCATCAACCCCTCTAACTGGTGTGATGGCGGTGCACACTGCC 531
 QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 532 GCGCGGAGAGCAGAGATCGGTGTTCGCTCTACGACCAACCTCATCTCCCTAGGTG 591
 QY 161 TyrSerSerGlnArgLysSerTyrPheProValCysGlnAspAspTyrAsnGluAsnTyr 180
 Db 592 TACTCATCTCAGAGAAAGTCCTGGCACCCTGTGTGCCAAGACGACTGGAACGAGAACTAC 651
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 Db 652 GGGCGGGCGGCTCAGGGACATGGGTATAGAAATAATTTTACTTACTACCCAGGATTA 711
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 712 GTGGATGACACGGATCCACAGCTTTATGAACCTGAACCAAGTCGGCGCAATGTGAT 771
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Db 772 ATCTATAAAAACTGTACACAGTATGCTCTCTTCAAAAGCGTGGTTCCTTACGC 831

QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 Db 832 TGTATAGCTCGGGGTCAACTTGAAGCTCAAGCCGAGAGCAGGATTTGGGGCGGAG 891
 QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 Db 892 AGCGGCTCCCGGGGCTGGCCCTGGCAGGTACGCTGCACGTCCAGAGCTCCAGGTG 951
 QY 281 CysGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
 Db 952 TCGGAGGCTCCATCATCACCCCGAGTGGATCGTACAGCGGCCCTCTGGTGGAAAAA 1011
 QY 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 Db 1012 CCTCTTAACAATCCATGGCAATTGGACGCAATTGGGGGATTTTGACACATCTTTATG 1071
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 Db 1072 TTCTATGAGCGGATACCAAGTAGAAAGATGATTCTCTCAAAATTTAGACTCCAAAG 1131
 QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 Db 1132 ACCAAGAACAATGATTCGCTGATGAGGTGAGAGCTCTGACTTTTCAACGACCTA 1191
 QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
 Db 1192 CTGAAACCCAGTGTCTGTGCCCAACCCAGCATGATGTCTGCAGCCAGAACAGCTTGTG 1251
 QY 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 Db 1252 ATTTCCGGGTGGGGGCGCCCGAGAGAAAGGAGACCTCGAAGTGTCTGACGCTGCC 1311
 QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 Db 1312 AAGTGTCTTCATTGAGACACAGATGTCAACAGCAGATATGTCTATGACAACTGATC 1371
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyVas 440
 Db 1372 ACACAGGACCATGATCTGTGGCGGCTTCTCGAGGGAAACGTCGATTTCTTGCAGGGTAC 1431
 QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrPheLeuIleGlyAspThrSer 460
 Db 1432 AGTGAGGGGCTCTGTGTCACTTCGAAGAACAAATATCTGGTGGCTCATAGGGGATACA 1491
 QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 Db 1492 TGGGTTCTGGCTGTGCCAAGCTTACAGACAGAGGTACGGAGTGTACGGATGTGTATTC 1551
 QY 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
 Db 1552 ACGGACTGGATTATCGACAAATGAGGGCAGACGCGC 1587

RESULT 4

AF270487

LOCUS

DEFINITION

AF270487.1

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

1740 bp mRNA linear PRI 04-APR-2001
 Homo sapiens androgen-regulated serine protease TMPRSS2 precursor
 (TMPRSS2) mRNA, complete cds.

AF270487
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1740)
 Afar, D.E., Vivanco, I., Hubert, R.S., Kuo, J., Chen, E., Saffran, D.C.,
 Raitano, A.B. and Jakubovits, A.
 Catalytic cleavage of the androgen-regulated TMPRSS2 protease
 results in its secretion by prostate and prostate cancer epithelia
 Cancer Res. 61 (4), 1686-1692 (2001)
 21139112
 11245484

REFERENCE 2 (bases 1 to 1740)
 AUTHORS Mitchell, S.C., Hubert, R.S. and Afar, D.E.H.
 TITLE Direct Submission
 JOURNAL Submitted (19-MAY-2000) UroGenesys, Inc., 1701 Colorado Ave., Santa Monica, CA 90404, USA

FEATURES
 source
 1..1740
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 1..1740
 /gene="TMPRSS2"
 114..1592
 /gene="TMPRSS2"
 /note="expressed in prostate and prostate cancer; differs from product encoded by the sequence presented in GenBank Accession Number U75329"
 /codon_start=1
 /product="androgen-regulated serine protease TMPRSS2 precursor"
 /protein_id="AAK29280.1"
 /db_xref="GI:13540004"
 /translation="MALNSGPPAIGPYENHGYQENPYPAQPTVPTVVEVHPAQY YSPVQAPVRLTQASNPVCTQPKSPGVCTSKTKKALCITLTGTFVUGAALAA GLLWFMGSKCSNGIECDSDGTCINPNWCDGVSHCPGGEDENRCLYGNFILLQV YSSQKSHPVQDDWNNYGRAARDMGYKNNFYSSQGIYDDSGSISFMKLTNSAGN VDIYKXLSHSDASKVAVSLRACIAGVNLNSROSRIYGGESALPGAWQVSLRVQ NVHYGGSIITPEWTVTAHCVKEPLNPNHWTAFAPIRQSFYFGAPYQVEVLIH PNYDSKTNKNDILMKLQKPLTFDNLVPVCLPNPMMLOPEQLCMTSGMGATEEKGK TSEVLNAAKLLIETQRNSRYVDNLITPAMICAGFLQGNVDSQGDGSGPLVTSKN NIWLIGDTSWGSQCAKAYRPGVYGNVNVFTDWIYRQWRADG"
 879..1589
 /gene="TMPRSS2"
 /product="androgen-regulated serine protease TMPRSS2"

mat_peptide
 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPhelleLeuGlnVal 160
 534 GCGGGGAGGAGCAGAAATCGTGTGCTCTACGACCAAACTTCATCTTCAGGTG 593
 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspSerTrpAsnGluAsnTyr 180
 594 TACTCATCTCAGAGGAAGTCTTGGCACCTCTGTGCAAGACGACTGGAAACGAACTAC 653
 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
 654 GCGGGGCGGCTGCAGGGACATGGGCTATAAGAAATAATTTTACTCTAGCAAGGAATA 713
 201 ValAspSerSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 714 GTGGATGACAGCGGATCCACCGCTTTATGAAACTGAACAAGTCCGCGCAATGTCGAT 773
 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 774 ATCTATATAAACTGTACCACTGATGCTGCTTCTTCAAAAGCAGTGGTTCCTTTACGC 833
 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 834 TGTATAGCTCGGGGTCAACTTGAACCTCAAGCGCCAGACGAGGATTTGGGCGCGGAG 893
 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
 894 AGCGCGCTCCCGGGGCTGGCCCTGGCAGTCAGCTGCACCTGCAGACGTCACGCTG 953
 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
 954 TCGGAGGCTCCATCATCATCCCGCGAGTGGATCGTGACAGCGCCCACTCGTGGAAAA 1013
 301 ProLeuAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 1014 CCTCTTAACCAATCCATGGCAATTTGACGCGCATTTTGGGGGATTTTGACACATCTTTCATG 1073
 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 1074 TTCTATGAGCGCGATACCAAGTAGAAAAAGTAGTATTCTCATCAAAATATAGTCCAAAG 1133
 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 1134 ACCAAGAACCAATGACATTCGCTGATGAGCTGAGAGCTCGAGAGCTCTGACTTTCAACGACCTA 1193
 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
 1194 GTGAACCACTGTGTCTGCCCAACCCAGGCGATGATGCTGCAGCCAGAACAGCTCTGCTGG 1253
 381 IleSerGlyTrpGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 1254 ATTTCCGGGTGGGGGCGCCACCGAGGAGAAAGGAAAGACCTCAGAGTGTCTGACAGCTGCC 1313
 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 1314 AAGTGTCTTCTATTGAGACACAGAGATGCAACAGCAGATATGCTATGACCAACCTGATC 1373
 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 1374 ACACCAAGCATGATCTGTGCGGCTTCCTGCAGGGGAACGCTCGATTTCTGCGAGGCTGAC 1433
 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
 1434 AGTGGAGGGCTCTGGTCACTTCGAGAACAAATATCTGGTGGCTGATAGGGGATACAGC 1493
 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 1494 TGGGGTCTGGTGTGCCAAAGCTTACAGACGAGGAGTACGGGAATGTGATGGTATTC 1553
 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
 1554 ACGGACTGGATTATCGACAAATGAGGCGACGCGC 1589

ORIGIN
 Alignment Scores:
 Pred. No.: 3,82e-197 Length: 1740
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-09-615-285B-2 (1-492) x AF270487 (1-1740)

QY 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Db 114 ATGGCTTTGAACCTCAGGGTCAACCCAGCTATTTGGACCTTACTATGAAACCACTGGATAC 173
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 Db 174 CACCGGAAACCCCTATCCGACACGCCCACTGTGGTCCCACTGTCTACGAGGTGCAT 233
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 234 CCGGCTCAGTACTACCGTCCCGGTGCCAGTACCGCCCGGAGGCTCTGACGCGAGCT 293
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 294 TCAACCCCGTGTCTGACGAGGAGCCCAATCCCAATCCGAGACAGTGTGACCTCAAG 353
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 Db 354 ACTAAGAACGACTGTGCATCACCTTGACCTGGGACCTTCTCTGTTGGAGCTCGGCTG 413
 QY 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 414 GCGGCTGGGCTACTCTGGAAGTTATGGGACAGAAAGTGTCCAACCTCTGGGATAGAGTGC 473
 QY 121 AspSerSerGlyThrCysIleLeuAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 Db 474 GACTCTCAGGTTACTTGCATCAACCCCTTACTGTTGATGGGCTGTCACTGCCCC 533

```
RESULT 5
BC051839
LOCUS
DEFINITION
  BC051839 3195 bp mRNA linear PRI 07-OCT-2003
  Homo sapiens transmembrane protease, serine 2, mRNA (cDNA clone
  HGC:60144 IMAGE:6199625), complete cds.
ACCESSION
  BC051839
VERSION
  BC051839.1 GI:30353994
KEYWORDS
  MGC.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  1 (bases 1 to 3195)
  Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
  Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
  Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
  Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
  Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
  Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
  Scheet, T.E., Brownstein, M.J., Usdin, T.B., Tothiyuki, S.,
  Canninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
  Abramson, R.D., Mullighy, S.J., Bosak, S.A., McEwan, P.J.,
  McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
  Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
  Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
  Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
  Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
  Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
  Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
  Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E.,
  Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
  Generation and initial analysis of more than 15,000 full-length
  human and mouse cDNA sequences
  Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  22398257
  12477932
  2 (bases 1 to 3195)
  Strausberg, R.
  Direct Submission
  Submitted (01-MAY-2003) National Institutes of Health, Mammalian
  Gene Collection (MGC), Cancer Genomics Office, National Cancer
  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
  USA
  NIH-MGC Project URL: http://mgc.nci.nih.gov
  Contact: MGC help desk
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: Dr. James R. Lupski
  CDNA Library Preparation: Life Technologies, Inc.
  DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Sequencing Group at the Stanford Human Genome
  Center, Stanford University School of Medicine, Stanford, CA 94305
  Web site: http://www-shgc.stanford.edu
  Contact: (Dickson, Mark) mcd@paxil.stanford.edu
  Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
  R. M.
  Clone distribution: MGC clone distribution information can be found
  through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
  Series: IRAK plate: 110 Row: 0 Column: 3
  This clone was selected for full length sequencing because it
  passed the following selection criteria: matched mRNA gi: 14602458.
  Location/Qualifiers
    1. .3195
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="MGC:60144 IMAGE:6199625"
      /tissue_type="Peripheral Nervous System, sympathetic
      trunk"
      /clone_lib="Lupski_sympathetic_trunk"
      /lab_host="DH103"
      /notes="Vector: pCMV-SPORT6"
    1. .3195
      gene
FEATURES
  source
    /gene="TMPRSS2"
    /note="synonym: PRSS10"
    /db_xref="LocusID:7113"
    /db_xref="MIM:602060"
    106_..1584
    /codon_start=1
    /product="transmembrane protease, serine 2"
    /protein_id="AAHS1839.1"
    /db_xref="GI:30353995"
    /db_xref="LocusID:7113"
    /translation="MALNSGPPAIGPYFYNHGVOPENPYPAQPTVPTVEVHPAQY
    YPSVPQAPRVLTQASNPVCTOPKSPGTVCTKTKKALCITLTGLTFLVGAALAA
    GLLWMFGSKCSNGIECDSSGTCINPNCWCDGVSHCPGDEENRCVRLYGFNFILQV
    YSSQKSHVPQCODDWNENYGRACRDVGNKNFYSSQIVDDSGSTFPMKLTAGN
    VDIYKLVHSDACSSKAVYLRCIACGVNLNSRSRIVGGESALPGAWPQVSVLHVQ
    NVHVGCGIIPFENIVTAHCEKELNNPHTWTAAGILROSFMFYGAGYQVEKVISH
    PYNDKNDNDIALMKLQKPLTFNDLVFVCLPNPQMQLQPEQLCMIWGWGATEEGK
    TSEVNLAAKLVLIETORCNSRYVYDNLITPAMICAGFLQGNVDSCQSGSGGLVTSKN
    NTWLLIGDTSWGGCAKAYPGVYGVNVVFTDWIYRQMRADG"
    550_..828
    misc_feature
      /note="SR; Region: Scavenger receptor Cys-rich"
      /db_xref="CDD:smart00202"
    868_..1557
    misc_feature
      /note="tryp_SPC; Region: Trypsin-like serine protease"
      /db_xref="CDD:smart00020"
ORIGIN
  Alignment Scores:
  Fred. No.: 7,87e-197 Length: 3195
  Score: 2717.00 Matches: 492
  Percent Similarity: 100.00% Conservative: 0
  Best Local Similarity: 100.00% Mismatches: 0
  Query Match: 100.00% Indels: 0
  DB: 9 Gaps: 0
  US-09-615-285b-2 (1-492) x BC051839 (1-3195)
  QY 1 MetAlaLeuAanSerGlySerProAlaIleGlyProTyrTyrGluAenHisGlyTyr 20
  Db 106 ATGGCTTTGAATCAGGTCACCCAGCAGTATTGGACCTTACTATGAAACCATGATAC 165
  QY 21 GlnProGluAanProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
  Db 166 CAACCCGGAAACCCCTATCCCGCAGCCACCTGTGTGTCCTCTACGAGGTGCAT 225
  QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
  Db 226 CCGGCTCAGTACTACCGCTCCCGCTCCCGCCAGTACGCCCGGAGGTCTCTGACGCGCT 285
  QY 61 SerAanProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
  Db 286 TCCAAACCCGCTGCTGCGACGCGAGCCAAATCCCATCCGGACACACTGTGCACCTCAAAG 345
  QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
  Db 346 ACTAAGAAAGCACTGTGCATCACTTGAACCTGGGACCTTCTCTGGGAGCTGCGCTG 405
  QY 101 AlaAlaGlyLeuLeuTyrPlyPheMetGlySerLysCysSerAanSerGlyIleGluCys 120
  Db 406 GCGCGTGGCTACTCTGGAGATTCTATGGGAGCAAGTGTCCCAACTCTGGGATAGAGTGC 465
  QY 121 AspSerSerGlyThrCysIleAanProSerAanTyrCysAspGlyValSerHisCysPro 140
  Db 466 GACTCTCAGTACCTGTGCATCAACCCCTCTTAACCTGGTGTGATGGCTGTGCACACTGCC 525
  QY 141 GlyGlyGluAspGluAanArgCysValArgLeuTyrGlyProAenPheIleLeuGlnVal 160
  Db 526 GCGCGGAGGACGAGAAATCGGTGTGTGCGCTCTACGGACCAAACTTCATCTTCAGGTG 585
  QY 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAanAspTyrAanGluAanTyr 180
  Db 586 TACTCACTCAGAGGAGAGTCTCTGGCACCCCTGTGTGCCAAGACGACTGGACGAGAACTAC 645
```


QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 646 GGG 705
QY 201 ValAspSerGlySerThrSerPheMetLysLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 706 GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACACAGTGGCGGCAATGTCGAT 765
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 766 ATCTATAAAAACTGTATACACAGTATCCCTGTTCTTCAAAAGCAGTGGTTCTTTACGC 825
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 826 TGTATAGCTCGGGGGCTGACCTTGAACCTCAAGCCGACAGCAGGATCGTGGCGCGAG 885
QY 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 886 AGCGCGCTCCGG 945
QY 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db 946 TCGGAGGCTCCATCATCACCCCGAGTGGATCGTGACAGCGCCGACCTGGTGGGAAAA 1005
QY 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 1006 CCTCTTAACATCCATGCGCATTTGGAGCGCAATTTGGGGGATTTTGACACATCTTTTCATG 1065
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAsnSerLys 340
Db 1066 TTCTATGAGCGCGGATACCAAGTGAAGAAAGTATTTCTCATCAATATGATCTCAAG 1125
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1126 ACCAAGAACAATGACATTCGGCTGATGAAGCTGCAGAGCCCTCTGACCTTCAACGACCTA 1185
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1186 GTGAACACCATGCTGCTGCCAACCCAGCGCATGCTGCAGCGGAGACACACCTCTGCTGG 1245
QY 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1246 ATTTCCGGGTGG 1305
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAsnLeuIle 420
Db 1306 AAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGCTATGACAACTGATC 1365
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1366 ACCAGCGCATGATCTGTCCCGGCTTCTTCAGGGGAAAGCTGATCTTTCAGCGGTGAC 1425
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1426 AGTGAGGGGCTCTGGTCACTTCGAAGAACAAATATCTGGTGCTGATAGGGGATACAGC 1485
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1486 TGGGGTCTGGCTGTGCCAAAGCTTACAGACAGAGGTGTACGGGAATGTGATGTATTC 1545
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1546 ACGGACTGATTTATCGACAATAGGGGCGAGCGGC 1581

RESULT 6
AF123453
LOCUS Homo sapiens transmembrane serine protease 2 (TMPRSS2) mRNA, complete cds.
DEFINITION
ACCESSION AF123453
VERSION AF123453.1 GI:5001723
KEYWORDS
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 1479)
TITLE Mutation analyses of 268 candidate genes in human tumor cell lines
JOURNAL Genomics 74 (3), 352-364 (2001)
MEDLINE 21309069
PUBMED 11414763
REFERENCE 2 (bases 1 to 1479)
AUTHORS Teng,D.H.F., Chen,Y., Lian,L., Ha,P.C., Tavtigian,S.V. and Wong,A.K.C.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-1999) Cancer Division, Myriad Genetics, Inc., 390 Wakara Way, Salt Lake City, UT 84108, USA
FEATURES
source
1..1479
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22"
1..1479
/gene="TMPRSS2"
1..1479
/genes="TMPRSS2"
1..1479
/genes="TMPRSS2"
/codon_start=1
/product="transmembrane serine protease 2"
/protein_id="AAD37117.1"
/db_xref="GI:5001724"
translation="NALNSGSPALGPYYENHGYOPENYPYPAQFTVVPVYVEHPAQY
YSPYPOYAPRVLTOASNPVCTOPKPSGTVCTSKTKKALCITLTLTGTLVLAALAA
GLLWPMGKSCNSGIECDSSGTCINPSNKGDSVGHCGGDEENRCVRLVGNFLLQM
YSSQKSWHPVQDDNENYGRACRDMGKKNFYSSQGIYDDSGTSFPMKLNLSAGN
VYIKLXHSADSCASVSLRACIAGNSRQSRISVGGSSALFPGAPWQVSLRHVQ
NVHVCSSGSIITPEWIVTAHCVEKPLNNPWHWTAFAGILRQSFYFGAGYQVEKVISH
PNYDSKNDIALMKLQKPLFNDLKVPCLPNPGMILQPOLCWIISGWGATEKGGK
TSEVLNAAKVLLIETORCNSRYVDNLITPAMICAGFLQGNVDSQCGDSGGPLVTSKN
NFWLLGDTWSGSGCAKVRPGVYGNVWFTDIYRQMRADG"

ORIGIN
Alignment Scores: 5,31e-197 Length: 1479
Pred. No.: 2714.00 Matches: 491
Score: 2714.00
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.80% Mismatches: 0
Query Match: 99.89% Indels: 0
DB: 9 Gaps: 0
US-09-615-285B-2 (1-492) x AF123453 (1-1479)
QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 1 ATGGCTTTGAACCTCAGGTGACCCAGCTATTGGACCTTACTATGAAAACCATGGATC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 61 CAACCGAAAAACCCCTATCCGACAGCCACCTGTGTCGCCACTGTCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACTACCCGTCCTCCCGGTCCTCCCGGTCCTCCCGGTCCTCCCGGTC 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCGCTCTCTGACGACGCCCAATCCCATCCGACAGTGTGCATCTCAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 241 ACTAAGAAGACATGTGCATCCTTGACCTGGGACCTTCTCTGTTGGAGCTGGCGCTG 300
QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120

301 GCCCTGGCCCTACTCTGGAAGTTCATGGGAGCAAGTGCTCCAACTCTGGGATAGAGTGC 360
121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
361 GACTCTCAGTACTGATCAACCCCTCTAACTGGTGTGATGCGGTGTACACTGCCCC 420
141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPhelLeuGlnVal 160
421 GCGCGGGAGACGAGAAATCGGTGTGTCTACGACCAAACTTCATCTCTCAGAT 480
161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
481 TACTCATCTCAGAGAGTCTGACACCTGTGTGCCAAGCAGCTGGAACGAGAACTAC 540
181 GlyArgAlaAlaCysArgAspMetGlyTyrIlysAsnAsnPhetYrSerSerGlnGlyIle 200
541 GCGCGGGCGGCTCAGGAGCATGGCTATAAGAAATAATTTTACTCTAGCCCAAGGATA 600
201 ValAspAspSerGlySerThrSerPheMetIlysLeuAsnThrSerAlaGlyAsnValAsp 220
601 GTGGATGACACGGGATCACCAGCTTTATGAATGAACAGTGGCGGCAATGTGAT 660
221 IleTyrIlysLeuTyrHisSerAspAlaCysSerSerIlysAlaValValSerLeuArg 240
661 ATCTATAAAAACTGACACAGTGTGCTGTTCTTCAAAGCAGTGGTTCCTTTACGC 720
241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
721 TGTATAGCTCGGGGTCACTTGAACCTCAAGCCGACAGAGTGGTGGGGGGTGG 780
261 SerAlaLeuProGlyAlaTyrProTyrProlnValSerLeuHisValGlnAsnValHisVal 280
781 AGCGGCTCCCGGGGCTCGCCCTGGCAGTGCAGCTGCAGCTCCAGACGCTCCACGTG 840
281 CysGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
841 TCGGAGGCTCATCATCACCCCGAGTGGATCGTGACAGCGCCCACTGCGTGGAAAAA 900
301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
901 CCTCTTAAATPCCATGGCATTTGAGCGGCATTTGCGGGATTTTGACACATCTTTCTATG 960
321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerIys 340
961 TTCTATGAGCGGATACCAAGTAGAAAAGTGTATTTCTCATCCAAATTTATGACTCCAAG 1020
341 ThrIlysAsnAsnAspIleAlaLeuMetIlysLeuGlnLysProLeuThrPheAsnAspLeu 360
1021 ACCAAGAACATGATGCGCTGATGAAGCTGACAGAGCTCTGACTTCAAGACCTA 1080
361 ValIysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
1081 GTGAAACCACTGTGTCTGCCCAACCCAGCATGATGCTGCAGCAGAACAGCTCTGCTGG 1140
381 IleSerGlyTyrGlyAlaThrCluGluLysGlyIysThrSerGluValLeuAsnAlaAla 400
1141 ATTTCCGGGTGGGGGCGCCACCGAGAGAAAGGAAAGCTCAGAAAGTCTGAAAGCTGCC 1200
401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
1201 AAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGCATATGTCTATGACAACTGATC 1260
421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
1261 ACACCAAGCCATGATCTGTCCCGGCTTCTGACAGGAAAGCGTGCATTTTGCAGGGTGAC 1320
441 SerGlyGlyProLeuValThrSerIlysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
1321 AGTGGAGGCGCTTGTGTCTCTTGAAGAAACATATCTGGTGGCTGATAGGGGATACAAGC 1380
461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
1381 TGGGGTCTGCTGCTGCCAAAGCTTACAGCAGGAGTGTACGGGAATGTGATGATTC 1440

QY 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
Db 1441 ACGGACTGGATTTATCGACAAATGAGGACACGGC 1476

RESULT 7
AF329454 3226 bp mRNA linear PRI 16-MAY-2001
LOCUS Homo sapiens epitheliasin (TMPRSS2) mRNA, complete cds.
DEFINITION AF329454
ACCESSION AF329454
VERSION AF329454.1 GI:14091027
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Jacquinet,E., Rao,N.V., Rao,G.V., Zhengming,W., Albertine,K.H. and Hoidal,J.R.
TITLE Cloning and characterization of the cDNA and gene for human epitheliasin
JOURNAL Eur. J. Biochem. 268 (9), 2687-2699 (2001)
MEDLINE 21232025
PUBMED 11322890
REFERENCE 2 (bases 1 to 3226)
AUTHORS Jacquinet,E., Rao,N.V., Rao,G.V., Wang,Z., Albertine,K.H. and Hoidal,J.R.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-2000) Pulmonary Medicine, University of Utah, 50N. Medical Dr., Salt Lake City, UT 84132, USA

FEATURES
source
1..3226
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
1..3226
/gene="TMPRSS2"
128..1606
/gene="TMPRSS2"
/note="proteinase; PRSS10"
/codon_start=1
/product="epitheliasin"
/protein_id="AAK3559.1"
/db_xref="GI:14091028"

Gene
CDS
128..1606
/translation="MALNSGPPAIGPYENHYQFENPYPAQPTVFTVVEVHPAQY
YSPVPQAPRVLTQASNPVCTQKPSGTCTKTKKALCITLTGLTFLVGAALAA
GLLWFMGSKSGIEDSGTCINPNCWCDGVSCFGEEDENCRVLYGNFILQM
YSQRKSHPCQDDWENYGRAACKDMYKNNFYSSQIGVDDSGTSPMKLNTSAGN
VDLYKLHSDACSSKAVVSLACVGNLNSRSRIVGGSALPGAWPQVSLHVQ
NVHVCSSITPEWITVAHCVEKPLNNPWHMTAFAGILRQSFMYGAGYQVEKVIH
PNYDSKNDIALMKLQKPLFNDLVFVCLPNPQMLQPSQLQWISGWGATEKKG
TSEVLNAKVLIIETQRCNSRYVDNLITTPAMICAGFLQGNVDSQGDSGGPLVTSKN
NIWLIGDTSWGSCKAYRPGVYGVNVMVFTDVIYRQMRADG"

ORIGIN
Alignment Scores: 2,28e-196 Length: 3226
Pred. No.: 2711.00 Matches: 490
Score: 100.00% Conservative: 2
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 99.59% Indels: 0
Query Match: 99.78% Gaps: 0
DB: 9

US-09-615-285B-2 (1-492) X AF329454 (1-3226)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 128 ATGGCTTGAACCTCAGGTCACCAACGAGCTATTGGACCTTACTATGAAACCAAGGATAC 187
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40

Db 188 CAACCGGAAACCCCTATCCCGCACAGCCCACTGTGGTCCCCACTGTCTACGAGGTGCAT 247
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 248 CCGGCTCAGTACTACCGTCCCGCGTCCCGCGTCCCGCGTCCCGCGTCCCGCGTCCCGCGT 307
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 308 TCCAAACCGCGTGTGTGTCAGCAGCCCAAAATCCCACTCCCGGACAGTGTGCACCTCAAG 367
 QY 81 ThrLysLysAlaLeuValCysThrLeuThrLeuThrLeuThrLeuThrLeuThrLeuThrLeu 100
 Db 368 ACTAAGAAACACCTGTGCATCCTGACCTGACCTGCGGACCTTCTCTGCGGAGCTGCGCTG 427
 QY 101 AlaAlaGlyLeuLeuThrLysPheMetGlySerLysCysSerAsnSerGlyLeuGluCys 120
 Db 428 CCGCGTGCCTTACTCTGGAAGTTCATGGGACAGCAAGTCTCCAACTCTGGGATAGATGC 487
 QY 121 AspSerSerGlyThrCysLeuAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 Db 488 GACTCCTCAGTACTGTGCATCAACCCCTCTAACTGGTGTGATGGCGTGTGCACCTGCCCC 547
 QY 141 GlyGlyLeuAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 548 GCGGGGAGGACGAGATCGTGTTCGCTCTACGACCAAACTTCATCTTCAGATG 607
 QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspTrpAsnGluAsnTyr 180
 Db 608 TACTCATCTCAGAGGAAGTCTGCGACCTGTGCGCAAGACGACTGGAACGAGAACTAC 667
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
 Db 668 GCGCGGGCGGCTCGAGGACATGGCTATGAAGATATTTTACTCTAGCCAGGATA 727
 QY 201 ValAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 728 GTGATGACAGCGGATCCACAGCTTTATGAACCTGAACACAGTGGCGGCAATGTGCGAT 787
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValSerLeuArg 240
 Db 788 ATCTATAAAACATGTACACAGTATGCCCTGTCTTCCAAAGCAGTGGTTCCTTACGC 847
 QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyLeu 260
 Db 848 TGTATAGCTCGGGGTCACTTGAACCTCAAGCCCGCACAGACAGGATTTGGCGGTGAG 907
 QY 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
 Db 908 AGCGCGCTCCCGGGGCGCTGCGCTGGCAGGTACAGCTGCACGTCACAGAGTCCACGTG 967
 QY 281 CysGlyGlySerIleIleThrProGlnTrpIleValThrAlaAlaHisCysValGluLys 300
 Db 968 TGGGAGAGCTCCATCATCACCCCGAGTGGATCGTGACAGCCGCCCTGCGTGGAAAAA 1027
 QY 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 Db 1028 CCTCTTAACTCCATGCGCATTTGGACGGCATTTGGGGGATTTTGAGACATCTTTTCATG 1087
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 Db 1088 TTCTATGAGCGCGGATACCAAGTAGAAAAGTATTTCTCATCCAAATTATGATCCAG 1147
 QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 Db 1148 ACCAAGAACATGACATTGCGCTGATGAAGCTGCAGAGAGCTCTGACTTTTCAAGAGCTA 1207
 QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGlnLeuCysTrp 380
 Db 1208 GTGAACACCATGTGTCTGCCCAACCCAGCATGATGCTGCAGCAGAGACACTCTGCTGG 1267
 QY 381 IleSerGlyTrpGlyValAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 Db 1268 ATTTTCGGGTGGGGGCCACCCAGGAGAAAGGAGAGACTCTCAGAGTGTCTGAACGCTGCC 1327

QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 Db 1328 AAGGTGCTTCTCATTTGAGACACAGAGATGCCAACAGACATATGTCTATGACACCTGATC 1387
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 Db 1388 ACACGAGCCATATCTGTGCGGCTTCTGTCAGGGAAGCTGATTCITGCGAGGTGAC 1447
 QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
 Db 1448 AGTGGAGGCGCTCTGGTCACTTCGAAGAACATACTCTGGTGGCTGATAGGGGATACAAGC 1507
 QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 Db 1508 TGGGGTCTTGGCTGTGCCAAAGCTTACAGCAGAGGTACGGGAATGTGATGTATTC 1567
 QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
 Db 1568 ACGGACTGGATTTATGCACAAATGAGGGCAGACGGC 1603

RESULT 8

AX201158 1476 bp DNA linear PAT 29-AUG-2001
 LOCUS Sequence 788 from Patent WO0151633.
 DEFINITION AX201158
 ACCESSION AX201158
 VERSION AX201158.1 GI:15390913
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
 Reed, S.G., Kalos, M.D., Fanger, G.R., Day, C.H., Retter, M.W.,
 Stolk, J.A., Skeiky, Y.A., Wang, A. and Meagher, M.J.
 TITLE Compositions and methods for the therapy and diagnosis of prostate
 cancer
 JOURNAL Patent: WO 0151633-A 788 19-JUL-2001;
 CORIXA CORPORATION (US)
 FEATURES
 source
 1..1476
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Alignment Scores:
 Pred. No.: 1,27e-196 Length: 1476
 Score: 2709.00 Matches: 491
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 99.71% Indels: 0
 DB: 6 Gaps: 0
 US-09-615-285B-2 (1-492) x AX201158 (1-1476)
 QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Db 1 ATGGCTTGAACCTCAGGTCACCCAGCTATTGGACCTTACTATGAAAACCATGGATAC 60
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 Db 61 CAACCGGAAAAACCCCTATCCCGCACAGCCCACTGTGGTCCCCACTCTCTACGAGGTGCAT 120
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 121 CCGGCTCAGTACTACCGTCCCCCGTGGCCCGCAGTACGCCCGGAGGTCTCTGACGCGGCT 180
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 181 TCCAAACCGCGTGTGTGTCAGCAGCCCAAAATCCCACTCCCGGACAGTGTGCACCTCAAG 240

QY 81 ThrIysIysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 241 ACTAAGAAACATGTCGATCCTGACCTCGGGACCTTCTCTGTTGGAGCTGGCG 300
QY 101 AlaAlaGlyLeuLeuTrpIysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCCGCTGGCCTACTCTGGAGATTCTATGGGAGCAAGTGTCTCAACTCTGGATAGAGTGC 360
QY 121 AspSerSerGlyThrCysIleLeuAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
Db 361 GACTCTCAGTACCTGATCAACCTCTTAACCTGTGTGATGTGCTGATGCTGACACTGCC 420
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTrpGlyProAsnPheIleLeuGlnVal 160
Db 421 GGGGGGAGACAGAGATCGGTGTGTCTCTACGGATCAAACTTCTCTCTCAGGTG 480
QY 161 TyrSerSerGlnArgIysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTrp 180
Db 481 TACTCATCTCAGAGGAAGTCTGGCACCCTGTGTGCAAGACACTGGAACGAGAACTAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTrpLysAsnAsnPheTrpSerGlnGlyIle 200
Db 541 GGGGGGCGCCTGACGGACATGGGCTATAGAAATAATTTTACTCTAGCCAGAGAA 600
QY 201 ValAspAspSerGlySerThrPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 601 GTGGATGACAGCGGATCCACAGCTTATGAACACTGAACACAGTAGTCCGGCAATGTCGAT 660
QY 221 IleTrpIysIysLeuTrpHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATAAAACTGTACCAAGATGATGCTGTCTTCAAAAGCATGGTTCCTTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerArgGlnSerArgIleValGlyGlu 260
Db 721 TGTATAGCTTGGGGGTCAACTTGAACCTCAAGCGCGCAGAGAGGATGTGGCGCGCAG 780
QY 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 781 AGCGCGCTCCGGGGCGCTGGCCTGGCAGGTGAGCTGCACGTCCAGAACGTCACGTG 840
QY 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db 841 TGGGAGCTCCATCATCACCCCGAGTGTGATGTCGACAGCCGCCACTCGGTGGA 900
QY 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTTAACAATCCATGGCATGACGCGCATTTTGGGGGATTTTGAGACAATCTTTTCATG 960
QY 321 PheTrpGlyAlaGlyTrpGlnValGluLysValIleSerHisProAsnTrpAspSerLys 340
Db 961 TTCTATGGAGCCGGATACCAAGTAGAAGAGTGTCTCTCATCCAAATATGACTCCAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACAATGACATTGGCTGATGAAGCTGCAGAAAGCTCTGACTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnCysTrp 380
Db 1081 GTGAACACAGTGTCTGTGCCAACCCAGGATGATGTCGAGCAGAACAGCTCTGCTGG 1140
QY 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCCGGTGGGGGCCACCCGAGAGAGAGGAGACCTCAGAAAGTGTGAACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTrpValTrpAspAsnLeuIle 420
Db 1201 AAGGTGCTTCTCATTTGAGACACAGATGCAACAGCAGATATGCTATGACAACTTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACAGCCATGATCTGTGCCGCTTCTGTGAGGGGAAAGCTGATCTTCTGCCAGGTGAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460

Db 1321 AGTGGAGGCGCTCTGGTCACTTCGAAGACAATATCTGGTGGCTGATAGGGATACAAG 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTrpGlyAsnValMetValPhe 480
Db 1381 TGGGGTTCTGCTGTGCAAAAGCTTACAGACCGAGGTGTACGGGAATGTGATGTTC 1440
QY 481 ThrAspTrpIleTrpArgGlnMetArgAlaAspGly 492
Db 1441 ACGACTGGATTTATCGACAAATGAGGGCAGACGGC 1476
RESULT 9
LOCUS AX267957 1476 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 931 from Patent WO0173032.
ACCESSION AX267957
VERSION AX267957.1 GI:16516542
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Eutelestomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
Kalos, M.D., Fanger, G.R., Retter, M.W., Stolk, J.A., Day, C.H.,
Vedvick, T.S., Carter, D., Li, S.X., Wang, A., Skeiky, Y.A., Hepler, W.T.
and Henderson, R.A.
TITLE Compositions and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0173032-A 931 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source 1.1476
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores: 1.27e-196 Length: 1476
Pred. No.: 491
Score: 2709.00 Matches:
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: Gaps: 6
US-09-615-285B-2 (1-492) x AX267957 (1-1476)
QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTrpGluAsnHisGlyTyr 20
Db 1 ATGGCTTTGAACCTCAGGGTCAACCCAGCTATTGGACCTTACTATGAAACCATGGATAC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTrpGluValHis 40
Db 61 CAACCGGAAACCCCTATCCCGCAGCCCACTGTGTGCCCACTGTCTACGAGGTGAT 120
QY 41 ProAlaGlnTrpTrpProSerProValProGlnTrpAlaProArgValLeuThrGlnAla 60
Db 121 CGGGCTCAGTACTACCCGTCGCCCGTCCCGAGTACGCCCGGAGGTCTCTGACGAGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCCGTGTGTGACGAGCCCAATCCCATCCGGACAGTGTGACCTCAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 241 ACTAAGAAGCACTGTGCATCACCTTGACCTGGGAGCTTCTCTGTTGGAGTGGCTG 300
QY 101 AlaAlaGlyLeuLeuTrpIysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCCGCTGGCCTACTCTGGAGATTCTATGGGAGCAAGTGTCTCAACTCTGGATAGAGTGC 360
QY 121 AspSerSerGlyThrCysIleLeuAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140

Db 361 GACTCTCAGTACCTGATCAACCCCTTAACCTGGTGTGATGGCTGTCACTGCCCC 420
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPhelleLeuGlnVal 160
Db 421 GCGGGGAGGACGAGATCGGTGTGGCTCTACGATCAAACTTCTCTCAGGTG 480
Qy 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
Db 481 TACTCATCTCAGAGAAAGTCTGGCACCCCTGTGTGCCAACAGCACTGGAAACGAGACTAC 540
Qy 181 GlyArgAlaAlaCysArgAspMetGlyTyrIysAsnAsnPhetYrSerGlnGlyIle 200
Db 541 GGGGGGGGGCTCGAGGACATGGCTATAAGAAATAATTTTACTCTAGCCCAAGAAATA 600
Qy 201 ValAspSerSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 601 GTGGATGACACGGATCCACAGCTTTATGAACCTGACACAGCTGCCGGCAATGTGAT 660
Qy 221 IleTyrIysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATAAAACCTGTACACAGTATGCTCTTCTTCAAAAGCAGTGGTCTCTTTACGC 720
Qy 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 721 TGTATAGCTCGGGGCTCACTTGAACCTCAGCGCCGACAGAGAGATTTGGCGCGAG 780
Qy 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnValHisVal 280
Db 781 AGCGCGTCCGGGGCTGGCCCTGAGCTGAGCTGACGCTCCAGAACCTCCACGCTG 840
Qy 281 CysGlyGlySerIleLeuThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
Db 841 TCGGAGGCTCCATCATCACCCCGAGTGTGATCGTACAGCCGCCACCTCGCTGGAAAA 900
Qy 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTTAACATCCATGCGATGGACGGCATTTGCGGGGATTTTGAGACAACTCTTCATG 960
Qy 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 961 TTCTATGAGCGCGATACCAAGTGAAGAAAGTATTTCTATCCAAATATGACCTCAAG 1020
Qy 341 ThrIysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACATGACATTTGGCTGATGAAGCTGACAGAGCCCTGACATTTCAACGACCTA 1080
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
Db 1081 GTGAACCACTGTGTCTGCCCAACCCAGGATGATGCTGCAGACAGACAGCTCTGCTGG 1140
Qy 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCCGGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAAAGTGTCTGAACGCTGCC 1200
Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1201 AAGTGCTTCTATTGAGACACAGATGATGATGATGATGATGATGATGATGATGATGATG 1260
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACGACCATGATCTGTGGCGCTTCTCTGAGGGGAAAGCTGATTTTGCCAGGGTGAC 1320
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTyrLeuIleGlyAspThrSer 460
Db 1321 AGTGGAGGGCTCTGGTCACTTCGAAGAACATATCTGGTGCTGATAGGGGATACAAAGC 1380
Qy 461 TrrGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGGTCTGGCTGTGCCAAAGCTTACACAGCAGGAGTGTACGGGAATGTGTGTATTC 1440
Qy 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492

Db 1441 ACGGACTGGATTATTCGCAAAATGAGGGCGACGGC 1476
RESULT 10
AX201157
LOCUS AX201157 1479 bp DNA Linear PAT 29-AUG-2001
DEFINITION Sequence 787 from Patent WO0151633.
ACCESSION AX201157
VERSION AX201157.1 GI:15390912
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
Reed, S.G., Kalos, M.D., Panger, G.R., Day, C.H., Retter, M.W.,
Stolk, J.A., Skeiky, I.A., Wang, A. and Meagher, M.J.
TITILE Compositions and Methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0151633-A 787 19-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1..1479
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores: 1..276-196 Length: 1479
Pred. No.: 2709.00 Matches: 491
Score: 2709.00
Percent Similarity: 99.80% Conservatives: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
Gaps: 0
US-09-615-285B-2 (1-492) x AX201157 (1-1479)
Qy 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 1 ATGGCTTTGAACCTCAGGGTCACACAGCTATTGACCTTACTATGAAACCATGGATAC 60
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
Db 61 CAACCGGAAACCCCTATCCGCACAGCCCACTGTGTGCCCTCCCTCTAGAGGTGAT 120
Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACTACCCGTCCTCCCGTGGCCAGTACGCCCGAGGCTCTGACGAGGCT 180
Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCGCTGTCTGCACGACGACCCAAATCCCATCCGGACAGTGTGCACCTCAAG 240
Qy 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 241 ACTAAGAAAGCACTGTGCATCACCTTGACCTCCCTGGGAGCTTCTCTGGTGGAGCTGCGGTG 300
Qy 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCGGCTGGCTACTCTGAAAGTTATCGGCAGCAAGTCTCCAACTCTGGGATAGAGTGC 360
Qy 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
Db 361 GACTCTCAGGTACTGTGATCAACCCCTTAACTGGTGTGATGGCTGTCACTTGCCTCCC 420
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPhelleLeuGlnVal 160
Db 421 GCGGGGAGGACGAGATCGGTGTGTTCCCTCTACGATCAAACTTCTCTCAGGTG 480
Qy 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
Db 481 TACTCATCTCAGAGAAAGTCTGGCACCCCTGTGTGCCAACAGCACTGGAAACGAGACTAC 540

QY 191 GlyAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
 Db 541 GGGCGGGCGGCTTGGAGGACATGGGCTATAAGATAATATTTTACTCTAGCAAGGAATA 600
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 601 GTGGATGACAGCGGATCCACAGCTTATGAACCTGAACACAGTGGCGGCAATGTGAT 660
 QY 221 IleTyrLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Db 661 ATCTATAAATACTGTACCACAGTGTGCTCTTCTTCAAAGCAGTGTGTTCTTTACGC 720
 QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyLeu 260
 Db 721 TGATATAGCTGGGGGTCACTTGAACCTCAAGCCGACAGACAGATGTGGGGGGCGAG 780
 QY 261 SerAlaLeuProGlyAlaTyrProThrGlnValSerLeuHisValGlnAsnValHisVal 280
 Db 781 AGCGCGCTCCCGGGGCGCTGGCCCTGGCAGGTGACCTGCACCTCCAGAACCTCCACGTG 840
 QY 281 CysGlyGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGlyLys 300
 Db 841 TGGGAGGCTCCATCATCATCAACCCCGAGTGGATCGTGACAGCGGCCACTGGGTGGAANA 900
 QY 301 ProLeuAsnAsnProThrHisThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 Db 901 CCTCTTAAACATCCATGGCATTTGGACGGCATTTGGCGGGATTTTGAGACAACTTTTCATG 960
 QY 321 PheTyrGlyAlaGlyTyrGlnValGlyLysValIleSerHisProAsnTyrAspSerLys 340
 Db 961 TTCTATGGAGCGGATACCAAGTAGAANAAGTATTTCTCATCCAAATTTAGCTCCAG 1020
 QY 341 ThrLysAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 Db 1021 ACCAAGAACATGATTCGCTGATGATGAGTGCAGAGAGCCCTCGACTTTCAACAGACCTA 1080
 QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
 Db 1081 GTGAACACAGTGTGCTGCCCAACCCAGCATGATGCTGCAGCAGAACAGCTCTGCTGG 1140
 QY 381 IleSerGlyTyrGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 Db 1141 ATTTCCGGTGGGGGGCCACCGAGGAGAAAGGAGAGCCCTCAGAAAGTGTCTGAACGCTGCC 1200
 QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 Db 1201 AAGGTGCTTCTCATTCAGACACAGAGATGCAACAGCAGATATGCTATGACAACTGATC 1260
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 Db 1261 ACACAGCCCATGATCTGTCCGGCTTCTGCAGGGGACGCTCGATTTCTGCCAGGGGTGAC 1320
 QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTyrLeuIleGlyAspThrSer 460
 Db 1321 AGTGAGGGCTCTGGTCACTTCGAAGAACAAATATCTGGTGGCTGATAGGGGATCAAGC 1380
 QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 Db 1381 TGGGGTCTGGCTGTGCCAAAGCTTACAGACAGAGGTGTACGGGAATGTGATGGTATTC 1440
 QY 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
 Db 1441 ACGGACTGGATTTATCGACAAATGAGGCGACAGCGGC 1476
 RESULT 11
 AX267956
 LOCUS
 DEFINITION
 Sequence 930 from Patent WO0173032.
 ACCESSION
 AX267956
 VERSION
 AX267956.1 GI:16516541
 KEYWORDS
 Homo sapiens (human)
 SOURCE

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Xu J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
 Kalos, M.D., Fanger, G.R., Retter, M.W., Stolk, J.A., Day, C.H.,
 Vedwick, T.S., Carter, D., Li, S.X., Wang, A., Skeiky, Y.A., Hepler, W.T.
 and Henderson, R.A.
 TITLE Compositions and methods for the therapy and diagnosis of prostate
 cancer
 JOURNAL Patent: WO 0173032-A 930 04-OCT-2001;
 CORIXA CORPORATION (US)
 FEATURES
 Location/Qualifiers
 source
 1..1479
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,27e-196 Length: 1479
 Score: 2709.00 Matches: 491
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 99.71% Indels: 0
 DB: 6 Gaps: 0
 US-09-615-285B-2 (1-492) x AX267956 (1-1479)
 QY 1 MetalAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Db 1 ATGGCTTTTGAACCTCAGGGTCAACACAGCTATTGGACCTTACTATGAAAACCATGGATAC 60
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 Db 61 CAACCGGAAACCCCTATCCCGCACAGCCCACTGTGTGCTCCCACTGTCTACGAGGTGCAT 120
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 121 CCGGCTCAGTACTACCCGCTCCCGCTCCCGACAGTACCCCGAGGCTCTGACGAGCT 180
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 181 TCCAAACCCGCTGCTGCACGACGCCAAATCCCACTCCCGGACAGTGTGCACCTCAAAG 240
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 Db 241 ACTAAGAAAGCACTGTGCATCACCTTGACCTGGGACCTTCTCTGGGAGCTGGCTG 300
 QY 101 AlalaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 301 GCGGCTGGCTTACTCTGAAAGTTCATGGGACGCAAGTGTCTCAACTCTGGGATAGAGTGC 360
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 Db 361 GACTCTCAGTACTCTGCATCAACCCCTCTAATGTGTGTGTGGCTGTGCACACTGCCCC 420
 QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 421 GCGCGGGAGGACGAGAAATCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
 QY 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
 Db 481 TACTCATCTCAGAGAGAGTCTTGGACCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 Db 541 GGGCGGGCGGCTGCAGGACATGGGCTATAAAGAAATAATTTTACTCTAGCCAAAGAAATA 600
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 601 GTGGATGACAGCGGATCCACAGCTTTATGAAACTGTGAACACAAAGTGCAGCAATGTTCGAT 660

Qy	221	IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg	240
Db	661	ATCTATAAAAAAAGCTGTACACAGTGTGCTGTCTTTCAAAGCAGTGGTTCTTTACGC	720
Qy	241	CysIleAlaCysGlyValAenLeuAenSerSerArgGlnSerArgIleValGlyGlyGlu	260
Db	721	TGTATAGCCTTGGCGGGTCAACTTGAATCTCAAGCGCCAGACAGATGTGGCGCGCGAG	780
Qy	261	SerAlaLeuProGlyAlaTirProTirGlnValSerLeuHisValGlnAenValHisVal	280
Db	781	AGCGCGCTCCGGGGGCTGGCCCTGGCAGGTGACCTGCACCTCCAGAAGCTCCACGTG	840
Qy	281	CysGlyClySerIleIleThrProGluTirPirIleValThrAlaAlaHisCysValGluLys	300
Db	841	TGGCGAGGCTCCATCATCACCCCGAGTGTGATCGTGACGCCGCCCACTCGGTGGAAAAA	900
Qy	301	ProLeuAsnAsnProTirHisTirThrAlaPheAlaGlyIleLeuArgGlnSerPheMet	320
Db	901	CCTCTTAACATCCATGCGCATTTGCAGCGCATTTTGGCGGGATTTTGACACAATCTTTCATG	960
Qy	321	PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProLeuTyrAspSerLys	340
Db	961	TTCTATCGAGCGGATACCAAGTAGAAAAAGTGATTTCTCATCCAAATTATAGATCCAAG	1020
Qy	341	ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu	360
Db	1021	ACCAAGAACAAATGACATGGCGTGATGAAGCTGCAGAAAGCCTCTGACTTTCAACGACCTA	1080
Qy	361	ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTirP	380
Db	1081	GTGAACACCATGTGTCTGCCCAACCCAGGCATGATGCTGCAGCCAGAACAGCTCTGCTGG	1140
Qy	381	IleSerGlyTirGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAenAlaAla	400
Db	1141	ATTTCGGGTGGGGGGCCACCGAGGAAAGGGAAGACCTCAGAAGTGCTGGAACGCTGCC	1200
Qy	401	LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAenLeuIle	420
Db	1201	AAAGTGCTTCTCATTTGACACACAGAGATGCAACAGCAGATATGTCTATGACACCTGATC	1260
Qy	421	ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAenValAepSerCysGlnGlyAsp	440
Db	1261	ACACCGCCATGATCTGTGGCGGCTTCTCTGCAGGGGAACGTCGATTTCTGCCAGGGTGAC	1320
Qy	441	SerGlyGlyProLeuValThrSerLysAsnAsnIleTirTirPirLeuIleGlyAspThrSer	460
Db	1321	AGTGGAGGGCCTCTGGTCACTTCGAAAGAACAAATATCTGGTGGCTGTATAGGGGATACAAGC	1380
Qy	461	TirPglySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAenValMetValPhe	480
Db	1381	TGGGGTCTTGGCTGTGCCAAAGCTTACAGACCAGAGTGTACCGGAATGTGTGTTATTC	1440
Qy	481	ThrAspTirPirIleTyrArgGlnMetArgAlaAspGly	492
Db	1441	ACGGACTGGATTTATTCGACAAATCAGGGCGACAGCGC	1476

RESULT	12
AX201156	
LOCUS	AX201156 3245 bp DNA linear PAT 29-AUG-2001
DEFINITION	Sequence 786 from Patent WO0151633.
ACCESSION	AX201156
VERSION	AX201156.1 GI:15390911
KEYWORDS	.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
Eukaryota;	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia;	Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y., Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Retter,M.W., Stolk,J.A., Skeiky,Y.A., Wang,A. and Meagher,W.J. Compositions and methods for the therapy and diagnosis of prostate cancer
TITLE	

JOURNAL Patent: WO 0151633-A 786 19-JUL-2001/
CORIXA CORPORATION (US)
FEATURES Location/Qualifiers
source 1..3245
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:		
Pred. No.:	3.25e-196	Length:
Score:	2709.00	Matches:
Percent Similarity:	99.80%	Conservative:
Best Local Similarity:	99.80%	Mismatches:
Query Match:	99.71%	Indels:
DB:	6	Gaps:
		3245
		491

US-09-615-285B-2 (1-492) x AX201156 (1-3245)

Qy	1	MetAlaLeuAenSerGlySerProProAlaIleGlyProTyrTyrClnAenHisGlyTyr	20
Db	145	ATGGTTTGAACCTCAGGGTCACACCAAGCTATTGGACCTTACTATGAAACCATGGATAC	204
Qy	21	GlnProGluAenProTyrProAlaGlnProThrValValProThrValTyrGluValHis	40
Db	205	CAACCGGAAAACCCCTATTCCGCGACAGCCCACTGTGTCCTCCCACTGTCTACAGGTGCAT	264
Qy	41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60
Db	265	CGGGCTCAGTACTACCCGTCGCCCGTACGCCCGAGGGTCTCTGACGAGGCT	324
Qy	61	SerAenProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys	80
Db	325	TCCAACCCCGTCGTGCGACGACCCCAATCCCACTCCGGACAGTGTGCACCTCAAG	384
Qy	81	ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu	100
Db	385	ACTAAGAAAGCACTGTGCATCACCTTCACCTGGGACCTTCCTCTGGGAGCTCGGTG	444
Qy	101	AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys	120
Db	445	CGCGTGCCTTACTCTGGAAAGTTCATGGGAGCAAGTGTCTCCAACTCTCGGATAGAGTGC	504
Qy	121	AspSerSerGlyThrCysIleAenProSerAsnTyrCysAspGlyValSerHisCysPro	140
Db	505	GACTCTCTCAGGTACTTCATCAACCCCTCTAACTGGTGTGATGGCGTGTACACTGCCCC	564
Qy	141	GlyGlyGluAaspGluAenArgCysValArgLeuTyrGlyProAenPheIleLeuGlnVal	160
Db	565	GGCGGGAGGACGAGAACTCGGTGTGTTCGCTCTACGGATCAAACTTCATCTCTTAGGG	624
Qy	161	TyrSerSerGlnArgLysSerTyrHisProValCysGlnAaspAspTyrAsnGluAsnTyr	180
Db	625	TACTCATCTCAGAGGAGTCTCTGCACCTGTGTGCCAGACGACTGGAAACGAGAACTAC	684
Qy	181	GlyArgAlaAlaCysArgAspMetGlyTyrLysAenAenPheTyrSerSerGlnGlyIle	200
Db	685	GGCGGGCGGCTGTCAGGGACATGGGCTATAAGAATAATTTTTTACTCTTAGCCCAAGGAATA	744
Qy	201	ValAaspAspSerGlySerThrSerPheMetLysLeuAenThrSerAlaGlyAenValAasp	220
Db	745	GTGGATGACAGCGGATCCACCGCTTTATGAACCTGAACTGAACCAAGTCCCGGCAATGTCTAT	804
Qy	221	IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg	240
Db	805	ATCTATAAAAAACTGTACCACAGTGATGCCTGTCTTCAAAAGCAGTGTGTTCTTTACGC	864
Qy	241	CysIleAlaCysGlyValAenLeuAenSerSerArgGlnSerArgIleValGlyGlyGlu	260
Db	865	TGTATAGCCTTCGGGGTCAACTTGAATCAAGCCGCCGACGACAGGATTGTGGCGCGGAG	924
Qy	261	SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAenValHisVal	280

QY 321 PheTyrGlyAlaGlyTyrGlnValGluLeuValIleSerHisProAsnTyrAspSerLys 340
Db 1105 TTCTATGAGCGGATACCAAGTAGAAAAAGTATTTCTCATCCAAATTATGACTCAAG 1164
QY 341 ThrLysAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1165 ACCAAGAACATGACATGGCTGATGAAGTCCAGAACCTTCTGACTTTCAACGACTA 1224
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
Db 1225 GTGAACACAGTGTCTGCGCAACCCAGGAGCATGATGCTGCAGCAGAACAGCTGTCTGG 1284
QY 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAla 400
Db 1285 ATTTCCGGCTGGGGGCCACCGAGAGAAAGGAGACCTCAGAGTGTGTAACCTGCC 1344
QY 401 LysValLeuLeuLeuGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1345 AAGGTGCTTCTCATTGAGACACAGATGCACAGCAGATATGCTATGACAACTGATC 1404
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnLysAsp 440
Db 1405 ACACCGACCATGATCTGTGCGGCTCTCGAGGAGACCTGCATCTTGCAGAGTGAC 1464
QY 441 SerGlyCysProLeuValThrSerLysAsnAsnIleTyrTyrLeuIleGlyAspThrSer 460
Db 1465 AGTGGAGGGGCTCTGTGCTCACTTCGAGAAACAATATCTGTGCTGATAGGGGATACAAGC 1524
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1525 TGGGGTCTGCTGTGCTCAAGCTACAGACAGGAGTGATCGGGAGATGATGATGATTC 1584
QY 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
Db 1585 ACGGACTGATTTATCGAATAATAGGGCAGACGGC 1620
RESULT 14
BD244778 3244 bp DNA linear PAT 17-JUL-2003
LOCUS Specific nucleic acid useful in diagnosis and treatment of
DEFINITION prostatic cancer and analysis of polypeptide expression.
ACCESSION BD244778
VERSION BD244778.1 GI:33054548
KEYWORDS JP 2002531061-A/3.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Macbeth,K.J. and Shyjan,A.W.
Specific nucleic acid useful in diagnosis and treatment of
prostatic cancer and analysis of polypeptide expression
Patent: JP 2002531061-A 3 24-SEP-2002;
MILLENNIUM PHARMACEUTICALS INC
OS Homo sapiens (human)
FN JP 2002531061-A/3
PD 24-SEP-2002
PF 30-SEP-1999 JP 2000572408
PR 30-SEP-1998 US 09/164159,30-SEP-1998 US 09/163759 PI
PC C12Q1/68,A61K45/00,A61K48/00,A61P35/00,C12Q1/02,G01N33/15, PC
G01N33/50,
PC G01N33/53,G01N33/53,G01N37/00/C12N15/09,(C12N15/09,C12R1:91),
PC (C12Q1/02,C12R1:91),C12N15/00,(C12N15/00,C12R1:91) CC
Specific nucleic acid useful in diagnosis and treatment of CC
prostatic
CC cancer and analysis of polypeptide expression FH key
Location/Qualifiers (145)...(1620).
FT CDS
1..3244 Location/Qualifiers
/organism="Homo sapiens"

/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores: 7.8e-196 Length: 3244
Pred. No.: 2704.00 Matches: 489
Percent Similarity: 99.80% Conservative: 2
Best Local Similarity: 99.39% Mismatches: 1
Query Match: 99.52% Indels: 0
DB: 6 Gaps: 0
US-09-615-285B-2 (1-492) x BD244778 (1-3244)
QY 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 145 ATGGCTTTTGAACCTCAGGGTCACACAGCTATTGGACCTTACTATGAAAACCATCGATAC 204
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 205 CAACCGGAAAACCCCTATCCCGACAGCCACTGTGGTCCCACTGTCTACAGGTGAT 264
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 265 CCGGCTCAGTACTACCCGCTCCCGCTGCCAGTACGCCCGGAGGCTCTGACGAGGCT 324
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 325 TCCAAACCCGCTGCTGTGACGCGAGCCCAATCCCATCCGGGACAGTGTGCACTCAAG 384
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 385 ACTAAGAAAGCACTGTGATCACCCTTGACCTCGGGACCTTCTCTGGGAGCTCGCGTG 444
QY 101 AlaAlaGlyLeuLeuTyrPhePheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 445 GCGCTGCGCTACTCTGGAAGTTCATGGGAGCAAGTGTCCAACTCTGGGATAGATGC 504
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
Db 505 GACTCTCAGGTACCTGCTCAACCCCTCTAACTGATGATGATGATGATGATGATGATGAT 564
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGluVal 160
Db 565 GCGGGGAGGAGGAGATCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 624
QY 161 TyrSerSerGlnArgLysSerTyrPheProValCysGlnAspAspTyrAsnGluAsnTyr 180
Db 625 TACTCATCTCAGAGGAGTCTCTGCGCAACCTGCTGCGCAACGAGTCTGGAACGAGACTAC 684
QY 181 GlyArgAlaAlaCysAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 685 GGGCGGGCGGCTCGAGGAGATGGCTATGAAGATTAATTTTACTCTGACCAAGGATA 744
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 745 GTGGATGACAGCGGATCCACGAGCTTTTGAACCTGAACTGAACTGAACTGAACTGAACT 804
QY 221 IleTyrLysValLeuTyrHisSerAspAlaCysSerSerLysValAlaValSerLeuArg 240
Db 805 ATCTATAAAAACTGTACCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 864
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 865 TGTTTAGCTCGGGGCTCACTTGAACCTCAAGCGCCAGAGGATCGTGGGCGGTGAG 924
QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
Db 925 AGCGGCTCCCGGGGCTGCGCTGCGAGTGCAGCTGCAGCTGCAGACAGCTCCAGCTG 984
QY 281 CysGlyGlySerIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
Db 985 TCGGAGGCTCCATCATCACCCCGAGTGGATGCTGACAGCCCGCCACTCGTGGTGAATA 1044

Qy 301 ProLeuAsnNanProTrrpHisTrrpThrAlaPheAlaGlyLeuArgGlnSerPheMet 320
 Db 1045 CCTCTTAACAAATCCATGGCAATTCGACGGCATTTGGGGGATTTTGAGCAATCTTTCATG 1104
 Qy 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 Db 1105 TTTCTATGGAGCGGATACCAAGTACAAAAGTATTTCTCATCCAAATATGACTCCAG 1164
 Qy 341 ThrLysAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 Db 1165 ACCAAGAACATGACATTCGCTGATGAAGCTGCAAGCCCTCTGACTTTCAACGACCTA 1224
 Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrrp 380
 Db 1225 GTGAACACAGTGTCTGCCCCAACCCAGCATGATGCTGCAGCCAGAACAGCTCTGCTGG 1284
 Qy 381 IleSerGlyTrrpGlyAlaThrGluGlnLysGlyLysThrSerGluValLeuAsnAlaLa 400
 Db 1285 ATTTCGGGTGGGGGGCCACCCAGGAGAAAGGAGACCTCAGAAAGTCTGAACGCTGCC 1344
 Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 Db 1345 AAGGTGCTTCTCATTCAGACACAGATGCAACACAGATATGCTATGACAACTGATC 1404
 Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 Db 1405 ACACCAAGCATGATCTGTCGGCTCTCTGCAGGGGAAAGCTGATTTCTTGCAGGGGTGAC 1464
 Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrrpTrrpLeuIleGlyAspThrSer 460
 Db 1465 AGTGGAGGGCTCTGCTGCTACTTCGAAGAACATATCTGGTGGCTGATAGGGGATACAGC 1524
 Qy 461 TrrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 Db 1525 TGGGGTCTGCGTGTGCCAAGCTTACAGACAGAGGTGTACGGGAAATGTGATGATTC 1584
 Qy 481 ThrAspTrrpIleTyrArgGlnMetArgAlaAspGly 492
 Db 1585 ACGGACTGGATTTATGCAAAATGAGGGCAGACGGC 1620

RESULT 15
 BD228306
 LOCUS
 DEFINITION
 Method of diagnosis, observation, staging, imaging and treatment of
 prostatic cancer.
 BD228306
 VERSION
 GI:33038076
 KEYWORDS
 JP 2002527758-A/8.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 Salceda, S., Recipon, H. and Caferkey, R.
 Method of diagnosis, observation, staging, imaging and treatment of
 prostatic cancer.
 TITLE
 Patent JP 2002527758-A 8 27-AUG-2002;
 JOURNAL
 DIADEXUS INC
 OS Homo sapiens (human)
 COMMENT
 PN JP 2002527758-A/8
 PD 27-AUG-2002
 PF 19-OCT-1999 JP 2000576884
 PR 60/104737
 PI SUSANA SALCEDA, HERVE RECIPON, ROBERT CAFFERKEY PC
 G01N33/574, A61K39/395, A61K49/00, A61K51/00, A61P35/00, PC
 C07K16/32,
 PC C12N15/09, C12Q1/68, G01N33/577, A61K49/02, C12N15/00 CC Method
 of diagnosis, observation, staging, imaging and CC
 treatment of
 CC prostatic cancer
 FH Key Location/Qualifiers
 FT source I. .3443

FT

FEATURES
source

Location/Qualifiers
 1. .3443
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 8,378-196 Length: 3443
 Score: 2704.00 Matches: 492
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 0
 Query Match: 99.53% Indels: 1
 DB: 6 Gaps: 0

US-09-615-285B-2 (1-492) x BD228306 (1-3443)

Qy 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Db 149 ATGGCTTTGAACCTCAGGGTCAACCACGACTATTGACCTTACTATGAAACCATGGATAC 208
 Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
 Db 209 CAACCGGAAACCCCTATCCCGCACAGCCACTGTGTGCCACTGTCTACGAGGTGCAT 288
 Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 269 CCGGCTCAGTACTACCGTCCCGGTGCCCGACGTACGCCCGAGGGTCTCTGACGAGGCT 328
 Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 329 TCCAAACCCCGTGTGTGCACGACGCCAAATCCCGGACAGTGTGCACCTCAAG 388
 Qy 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 Db 389 ACTAAGAAGACACTGTGCATCACCTTGACCTGGGACCTTCTCTCGTGGAGCTGCCTG 448
 Qy 101 AlaAlaGlyLeuLeuTrrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 449 GCGGTGCGCTTCTGGAAGTTTCATGGCAGCAAGTGTCCCACTCTGGGATAGATGC 508
 Qy 121 AspSerSerGlyThrCysIleAsnProSerAsnTrrpCysAspGlyValSerHisCysPro 140
 Db 509 GACTCCTCAGGTACTGTGCATCAACCCCTCACTAAGTGTGTGATGCGGTGCACACTGCC 568
 Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTrrpGlyProAsnPheIleLeuGlnVal 160
 Db 569 GCGGGGAGGACGAGATCGGTGTTCGCTCTACGACCAAACTTCATCCITCAGGTG 628
 Qy 161 TyrSerSerGlnArgLysSerTrrpHisProValCysGlnAspAspTrrpAsnGluAsnTyr 180
 Db 629 TACTCATCTCAGAGGAAGTCTCGGCACCCCTGTGTGCCAAGCAGACTGGAAACGAGACTAC 688
 Qy 181 GlyArgAlaAlaCysArgAspMetGlyTrrpLysAsnAsnPheTyrSerSerGlnGlyIle 200
 Db 689 GCGCGGGGGCGCTCGAGGACATGGGTATTAAGAAATAATTTTACTCTACCAAGGAATA 748
 Qy 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 749 GTGGATGACACGGGATCCACCGAGCTTTATGAAACTGAACAAGTGCCTGCAATGTTCAT 808
 Qy 221 IleTrrpLysLysLeuTrrpHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Db 809 ATCTATAAAACATGTACACAGATGCTGTCTTCAAAAGACAGAGGTGTTCTTTACGC 868
 Qy 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 Db 869 TGTATAGCTCGCGGGTCAACTTGAACTCAAGCGCCGACAGCAGGATCGTGGCGCGCGAG 928
 Qy 261 SerAlaLeuProGlyAlaTrrpTrrpGlnValSerLeuHisValGlnAsnValHisVa 280
 Db 929 AGCGGCTTCCCGGGGCGCTCGCCCTGGGACAGTGCCTGCACGTCCAGAACGTCCACGT 988

QY 280 lCysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLy 300
Db |||||
989 GTGCGAGGCTCCATCATCACCCCGAGTGGTGTGACGCGCCCACTGCTGGAAAA 1048
QY 300 sProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMe 320
Db |||||
1049 ACCTCTTAACAATCCATGSCATTGGACGGCATTTCGGGGATTTTGACACAATCTTTCAT 1108
QY 320 tPheTyrglyAlaGlyTyrglnValGluLyValIleSerHisProAsnTyrglnSerLy 340
Db |||||
1109 GTTCTATGGAGCGGATACCAAGTAGAAAAAGTAGATTCTCATCCAAATTATGACTCCAA 1168
QY 340 sThrLyAsnAsnAspIleAlaLeuMetLyLeuGlnLyProLeuThrPheAsnAspLe 360
Db |||||
1169 GACCAAGACATGATGCTGCTGATGAGCTGCAGAGCCTCTGACTTTCACGACCT 1228
QY 360 uValLyProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTr 380
Db |||||
1229 AGTGAACCAAGTGTCTCTGCCAACCCAGGCATGATGCTGCAGCCAGAACAGCTCTGCTG 1288
QY 380 pIleSerGlyTrpGlyAlaThrGluGluLyGlyLyThrSerGluValLeuAsnAlaAl 400
Db |||||
1289 GATTTCCGGGTGGGGGCCACCGAGGAGAAAGGGAGACCTCAGAAAGTGTGACGCTGC 1348
QY 400 aLyValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrglnValTyrglnAsnLeuIl 420
Db |||||
1349 CAAGGTGCTTCTCATTGACACACAGAGATGCAACACAGATATGTCTATGACCACTCAT 1408
QY 420 eThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAs 440
Db |||||
1409 CACACAGCCATGATCTGTGCCGGCTTCTGCAGGGGAACGTCGATTCTTGCCAGGGTGA 1468
QY 440 pSerGlyGlyProLeuValThrSerLyAsnAsnIleTrpTrpLeuIleGlyAspThrSe 460
Db |||||
1469 CAGTGGAGGGCTCTGGTCACCTCGAGAGACATATCTGGTGGCTGATAGGGGATACAAG 1528
QY 460 rTrpGlySerGlyCysAlaLyAlaTyrglnArgProGlyValTyrglyAsnValMetValph 480
Db |||||
1529 CTGGGGTTCTGGCTGTGCCAAAGCTTACAGACCGAGGTACGGGAATGTGATGTATT 1588
QY 480 eThrAspTrpIleTyrglnMetArgAlaAspGly 492
Db |||||
1589 CACGACTGGATTATCGACAAATGAGGGCAGACGGC 1625

Search completed: June 1, 2004, 16:37:10
Job time : 5898 secs


```
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 586
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-295-027-586

Query Match      100.0%; Score 2717; DB 15; Length 492;
Best Local Similarity 100.0%; Pred. No. 5.1e-230;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNSGPPAIGYENHGYQENPYPAQPTVTVYEVHPAQYPSVPQYAPRVLTQA 60
DB 1 MALNSGPPAIGYENHGYQENPYPAQPTVTVYEVHPAQYPSVPQYAPRVLTQA 60
QY 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLWKFMSKCSNSGIEC 120
DB 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLWKFMSKCSNSGIEC 120
QY 121 DSSGTCINPSNWCDSVSHCPGGEDENRCVRLYGFNFILQVYSSQKSWHPVQDDNNENY 180
DB 121 DSSGTCINPSNWCDSVSHCPGGEDENRCVRLYGFNFILQVYSSQKSWHPVQDDNNENY 180
QY 181 GRAACRDMGYKNFYSSQGIYVDDSGTSMKLNSTAGNVDIYKLYHSDACSSKAVVSLR 240
DB 181 GRAACRDMGYKNFYSSQGIYVDDSGTSMKLNSTAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGGESALPGAWPQVSLHVQNVHVCVCGSIITPEWITVAACHVEK 300
DB 241 CIACGVNLNSRQSRIVGGESALPGAWPQVSLHVQNVHVCVCGSIITPEWITVAACHVEK 300
QY 301 PLNPNHWTAFAGILRQSFYAGYQVEKVIHSPNYDSKTNNDIALMKLOKPLTFNDL 360
DB 301 PLNPNHWTAFAGILRQSFYAGYQVEKVIHSPNYDSKTNNDIALMKLOKPLTFNDL 360
QY 361 VKPCLPNPQWMLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
DB 361 VKPCLPNPQWMLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLQGNVDSQCGSGGLPVTSKNNIWWLIGDTSWGSCKAKAYRPGYGNVWF 480
DB 421 TPAMICAGFLQGNVDSQCGSGGLPVTSKNNIWWLIGDTSWGSCKAKAYRPGYGNVWF 480
QY 481 TDWIYQWRADG 492
DB 481 TDWIYQWRADG 492

RESULT 2
US-10-205-823-415
; Sequence 415, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Heersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsley, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205.823
```

```
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 415
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-205-823-415

Query Match      99.8%; Score 2711; DB 14; Length 492;
Best Local Similarity 99.6%; Pred. No. 1.7e-229;
Matches 490; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNSGPPAIGYENHGYQENPYPAQPTVTVYEVHPAQYPSVPQYAPRVLTQA 60
DB 1 MALNSGPPAIGYENHGYQENPYPAQPTVTVYEVHPAQYPSVPQYAPRVLTQA 60
QY 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLWKFMSKCSNSGIEC 120
DB 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLWKFMSKCSNSGIEC 120
QY 121 DSSGTCINPSNWCDSVSHCPGGEDENRCVRLYGFNFILQVYSSQKSWHPVQDDNNENY 180
DB 121 DSSGTCINPSNWCDSVSHCPGGEDENRCVRLYGFNFILQVYSSQKSWHPVQDDNNENY 180
QY 181 GRAACRDMGYKNFYSSQGIYVDDSGTSMKLNSTAGNVDIYKLYHSDACSSKAVVSLR 240
DB 181 GRAACRDMGYKNFYSSQGIYVDDSGTSMKLNSTAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGGESALPGAWPQVSLHVQNVHVCVCGSIITPEWITVAACHVEK 300
DB 241 CIACGVNLNSRQSRIVGGESALPGAWPQVSLHVQNVHVCVCGSIITPEWITVAACHVEK 300
QY 301 PLNPNHWTAFAGILRQSFYAGYQVEKVIHSPNYDSKTNNDIALMKLOKPLTFNDL 360
DB 301 PLNPNHWTAFAGILRQSFYAGYQVEKVIHSPNYDSKTNNDIALMKLOKPLTFNDL 360
QY 361 VKPCLPNPQWMLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
DB 361 VKPCLPNPQWMLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLQGNVDSQCGSGGLPVTSKNNIWWLIGDTSWGSCKAKAYRPGYGNVWF 480
DB 421 TPAMICAGFLQGNVDSQCGSGGLPVTSKNNIWWLIGDTSWGSCKAKAYRPGYGNVWF 480
QY 481 TDWIYQWRADG 492
DB 481 TDWIYQWRADG 492

RESULT 3
US-09-759-143-932
; Sequence 932, Application US/09759143
; Patent No. US200202248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
```

APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C23
CURRENT APPLICATION NUMBER: US/09/759,143
CURRENT FILING DATE: 2001-01-12
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 932
TYPE: PRT
LENGTH: 492
ORGANISM: Homo sapiens
US-09-759-143-932

Query Match 99.7%; Score 2709; DB 9; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.6e-229;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MALNSGSPALPGYENHGYQENPYPAQPTVTVTVVEVHPAQYYPSPVQYAPRVLTQA 60
Db 1 MALNSGSPALPGYENHGYQENPYPAQPTVTVTVVEVHPAQYYPSPVQYAPRVLTQA 60
Qy 61 SNPVVCTQPKSPSGTCTKTKKALCITLTGLTFLVGAALAAAGLLWKFMSKCSNSGIEC 120
Db 61 SNPVVCTQPKSPSGTCTKTKKALCITLTGLTFLVGAALAAAGLLWKFMSKCSNSGIEC 120
Qy 121 DSSGTCINPNCWCDGVSHCPGGEDENRCVRLYGNSFILQVYSSORKSWHPVCQDDWNEY 180
Db 121 DSSGTCINPNCWCDGVSHCPGGEDENRCVRLYGNSFILQVYSSORKSWHPVCQDDWNEY 180
Qy 181 GRAACRDMGYKNFYSSQGIIVDDSGSTFSFKMLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNFYSSQGIIVDDSGSTFSFKMLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Qy 241 CIACGVNLSSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLSSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
Qy 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
Qy 361 VKPVCPLNPGMWLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLIIETQRCNSRYVDNLI 420
Db 361 VKPVCPLNPGMWLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLIIETQRCNSRYVDNLI 420
Qy 421 TPAMI CAGFLQGNVDS COGDSGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNWVF 480
Db 421 TPAMI CAGFLQGNVDS COGDSGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNWVF 480
Qy 481 TDWIYQMRADG 492
Db 481 TDWIYQMRADG 492

RESULT 4
US-09-780-669-932
Sequence 932, Application US/09780669
Patent No. US20020051977A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqui

APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C24
CURRENT APPLICATION NUMBER: US/09/780,669
CURRENT FILING DATE: 2001-02-09
NUMBER OF SEQ ID NOS: 943
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 932
TYPE: PRT
LENGTH: 492
ORGANISM: Homo sapiens
US-09-780-669-932

Query Match 99.7%; Score 2709; DB 9; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.6e-229;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MALNSGSPALPGYENHGYQENPYPAQPTVTVTVVEVHPAQYYPSPVQYAPRVLTQA 60
Db 1 MALNSGSPALPGYENHGYQENPYPAQPTVTVTVVEVHPAQYYPSPVQYAPRVLTQA 60
Qy 61 SNPVVCTQPKSPSGTCTKTKKALCITLTGLTFLVGAALAAAGLLWKFMSKCSNSGIEC 120
Db 61 SNPVVCTQPKSPSGTCTKTKKALCITLTGLTFLVGAALAAAGLLWKFMSKCSNSGIEC 120
Qy 121 DSSGTCINPNCWCDGVSHCPGGEDENRCVRLYGNSFILQVYSSORKSWHPVCQDDWNEY 180
Db 121 DSSGTCINPNCWCDGVSHCPGGEDENRCVRLYGNSFILQVYSSORKSWHPVCQDDWNEY 180
Qy 181 GRAACRDMGYKNFYSSQGIIVDDSGSTFSFKMLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNFYSSQGIIVDDSGSTFSFKMLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Qy 241 CIACGVNLSSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLSSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
Qy 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
Qy 361 VKPVCPLNPGMWLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLIIETQRCNSRYVDNLI 420
Db 361 VKPVCPLNPGMWLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLIIETQRCNSRYVDNLI 420
Qy 421 TPAMI CAGFLQGNVDS COGDSGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNWVF 480
Db 421 TPAMI CAGFLQGNVDS COGDSGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNWVF 480
Qy 481 TDWIYQMRADG 492
Db 481 TDWIYQMRADG 492

RESULT 5
US-09-822-827-932
Sequence 932, Application US/09822827
Patent No. US20020081680A1

GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C1
; CURRENT APPLICATION NUMBER: US/09/822,827
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-822-827-932

Query Match 99.7%; Score 2709; DB 9; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.6e-229;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALNSGSPAIGPYENHGYQENPYPAQPTVVTVEVHPAQYPSVPOYAPRVLTQA 60
DB 1 MALNSGSPAIGPYENHGYQENPYPAQPTVVTVEVHPAQYPSVPOYAPRVLTQA 60
QY 61 SNPVCTOPKPSGTVCTSKTKKALCITLTGTFVLGAAALAGLLKFMGSKCNSGIEC 120
DB 61 SNPVCTOPKPSGTVCTSKTKKALCITLTGTFVLGAAALAGLLKFMGSKCNSGIEC 120
QY 121 DSSGTCINPNWCDGVSHCPGEGEDENRCVRLYGNPFILOVYSSQKSHHPVCQDDWNEY 180
DB 121 DSSGTCINPNWCDGVSHCPGEGEDENRCVRLYGNPFILOVYSSQKSHHPVCQDDWNEY 180
QY 181 GRAACDMGYKNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
DB 181 GRAACDMGYKNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGSGESALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
DB 241 CIACGVNLNSRQSRIVGSGESALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
QY 301 PLNPNPWHWTAFAGILRQSFMYGAGYQVEKVIHPNYDSKTKNDIALMKLQKPLTFNDL 360
DB 301 PLNPNPWHWTAFAGILRQSFMYGAGYQVEKVIHPNYDSKTKNDIALMKLQKPLTFNDL 360
QY 361 VKPVCPLPNPQWMLQPEQLCWISGWGATEKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
DB 361 VKPVCPLPNPQWMLQPEQLCWISGWGATEKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLOQNVDSQCGSGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGYGVNVWF 480
DB 421 TPAMICAGFLOQNVDSQCGSGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGYGVNVWF 480

RESULT 6
US-09-895-793-932
; Sequence 932, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.

APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Basols, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Fanger, Gary R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.534C2
CURRENT APPLICATION NUMBER: US/09/895,793
CURRENT FILING DATE: 2001-06-29
NUMBER OF SEQ ID NOS: 982
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 932
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-895-793-932

Query Match 99.7%; Score 2709; DB 9; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.6e-229;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALNSGSPAIGPYENHGYQENPYPAQPTVVTVEVHPAQYPSVPOYAPRVLTQA 60
DB 1 MALNSGSPAIGPYENHGYQENPYPAQPTVVTVEVHPAQYPSVPOYAPRVLTQA 60
QY 61 SNPVCTOPKPSGTVCTSKTKKALCITLTGTFVLGAAALAGLLKFMGSKCNSGIEC 120
DB 61 SNPVCTOPKPSGTVCTSKTKKALCITLTGTFVLGAAALAGLLKFMGSKCNSGIEC 120
QY 121 DSSGTCINPNWCDGVSHCPGEGEDENRCVRLYGNPFILOVYSSQKSHHPVCQDDWNEY 180
DB 121 DSSGTCINPNWCDGVSHCPGEGEDENRCVRLYGNPFILOVYSSQKSHHPVCQDDWNEY 180
QY 181 GRAACDMGYKNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
DB 181 GRAACDMGYKNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGSGESALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
DB 241 CIACGVNLNSRQSRIVGSGESALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
QY 301 PLNPNPWHWTAFAGILRQSFMYGAGYQVEKVIHPNYDSKTKNDIALMKLQKPLTFNDL 360
DB 301 PLNPNPWHWTAFAGILRQSFMYGAGYQVEKVIHPNYDSKTKNDIALMKLQKPLTFNDL 360
QY 361 VKPVCPLPNPQWMLQPEQLCWISGWGATEKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
DB 361 VKPVCPLPNPQWMLQPEQLCWISGWGATEKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLOQNVDSQCGSGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGYGVNVWF 480
DB 421 TPAMICAGFLOQNVDSQCGSGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGYGVNVWF 480
QY 481 TDWIYRQMRADG 492
DB 481 TDWIYRQMRADG 492

RESULT 7
US-09-895-814-932
; Sequence 932, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.

```

; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-814-932

Query Match          99.7%; Score 2709; DB 9; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.6e-229;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALNSGSPPAIGPYENHGYOPENPYPAQPTVTVVEVHPAQYYPSPVQYAPRVLTOA 60
Db 1 MALNSGSPPAIGPYENHGYOPENPYPAQPTVTVVEVHPAQYYPSPVQYAPRVLTOA 60
QY 61 SNPVVCTQPKSPGTVCTSKTKKALCITLTGTLVGAALAAAGLLWKFMSKCSNSGIEC 120
Db 61 SNPVVCTQPKSPGTVCTSKTKKALCITLTGTLVGAALAAAGLLWKFMSKCSNSGIEC 120
QY 121 DSSGTCINPNSWCDGVSHCPGDEENRCVRLYGNFNLQVYSSORKSWHPVCQDDWNNY 180
Db 121 DSSGTCINPNSWCDGVSHCPGDEENRCVRLYGNFNLQVYSSORKSWHPVCQDDWNNY 180
QY 181 GRAACRDMGYKKNFYSSGGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAVVSLR 240
Db 181 GRAACRDMGYKKNFYSSGGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCGGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCGGSIITPEWIVTAAHCVEK 300
QY 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVISHPNYDSKTNNDIAMKLOKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVISHPNYDSKTNNDIAMKLOKPLTFNDL 360
QY 361 VKPVCPLNPGMMLQPEQLCWISGWBATEEKGKTSSEVLNAKVLIIETQRCNSRYVDNLI 420
Db 361 VKPVCPLNPGMMLQPEQLCWISGWBATEEKGKTSSEVLNAKVLIIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLOQVNDSCGGSGPLVTSKNNIWWLIGDTSWGSCKAYRPGYGNWVVF 480
Db 421 TPAMICAGFLOQVNDSCGGSGPLVTSKNNIWWLIGDTSWGSCKAYRPGYGNWVVF 480
QY 481 TDWIYRQMRADG 492
Db 481 TDWIYRQMRADG 492

```

RESULT 8

```

US-10-012-896-932
; Sequence 932, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yudiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Mantanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012,896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-932

```

```

Query Match          99.7%; Score 2709; DB 13; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.6e-229;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALNSGSPPAIGPYENHGYOPENPYPAQPTVTVVEVHPAQYYPSPVQYAPRVLTOA 60
Db 1 MALNSGSPPAIGPYENHGYOPENPYPAQPTVTVVEVHPAQYYPSPVQYAPRVLTOA 60
QY 61 SNPVVCTQPKSPGTVCTSKTKKALCITLTGTLVGAALAAAGLLWKFMSKCSNSGIEC 120
Db 61 SNPVVCTQPKSPGTVCTSKTKKALCITLTGTLVGAALAAAGLLWKFMSKCSNSGIEC 120
QY 121 DSSGTCINPNSWCDGVSHCPGDEENRCVRLYGNFNLQVYSSORKSWHPVCQDDWNNY 180
Db 121 DSSGTCINPNSWCDGVSHCPGDEENRCVRLYGNFNLQVYSSORKSWHPVCQDDWNNY 180
QY 181 GRAACRDMGYKKNFYSSGGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAVVSLR 240
Db 181 GRAACRDMGYKKNFYSSGGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCGGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCGGSIITPEWIVTAAHCVEK 300
QY 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVISHPNYDSKTNNDIAMKLOKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVISHPNYDSKTNNDIAMKLOKPLTFNDL 360
QY 361 VKPVCPLNPGMMLQPEQLCWISGWBATEEKGKTSSEVLNAKVLIIETQRCNSRYVDNLI 420
Db 361 VKPVCPLNPGMMLQPEQLCWISGWBATEEKGKTSSEVLNAKVLIIETQRCNSRYVDNLI 420

```


QY 421 TPAMICAGFLQGNVDSQGDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVWF 480
Db 421 TPAMICAGFLQGNVDSQGDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVWF 480
QY 481 TDWIYRQMRADG 492
Db 481 TDWIYRQMRADG 492

RESULT 9

US-10-144-678A-932
; Sequence 932, Application US/10144678A
; Publication No. US20030157089A1

GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillion, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Hepler, William T.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals y de Bassols, Carlota
; APPLICANT: Foy, Teresa M.
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Deng, Ta
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C28
; CURRENT APPLICATION NUMBER: US/10/144,678A
; CURRENT FILING DATE: 2002-08-12
; NUMBER OF SEQ ID NOS: 1033
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-678A-932

Query Match 99.7%; Score 2709; DB 14; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.6e-229;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVVEVHPAQYYPSPVQYAPRVLTOA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVVEVHPAQYYPSPVQYAPRVLTOA 60
QY 61 SNPVVCTOPKSPGTVCTSKTKALCITLTGTLVGAALAAAGLLWKPMGSKCSNSGIEC 120
Db 61 SNPVVCTOPKSPGTVCTSKTKALCITLTGTLVGAALAAAGLLWKPMGSKCSNSGIEC 120
QY 121 DSSGTCINPSNWCDCGVSHCPGDEENRCVRLYGNSFNLQVYSSQKSWHPVCQDDWNNY 180
Db 121 DSSGTCINPSNWCDCGVSHCPGDEENRCVRLYGNSFNLQVYSSQKSWHPVCQDDWNNY 180
QY 181 GRAACRDGMGYNKXNFYSSQGIYDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDGMGYNKXNFYSSQGIYDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLSSRSQRIYVGGESALPGAWPQVSLHVQNWHVCGGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLSSRSQRIYVGGESALPGAWPQVSLHVQNWHVCGGSIITPEWIVTAAHCVEK 300

Db 241 CIACGVNLSSRSQRIYVGGESALPGAWPQVSLHVQNWHVCGGSIITPEWIVTAAHCVEK 300
QY 301 PLNNPWHMTAFAGILRQSFYFAGYQVEKVI SHPNYDSKTKNNDIAMKLOKPLTFNDL 360
Db 301 PLNNPWHMTAFAGILRQSFYFAGYQVEKVI SHPNYDSKTKNNDIAMKLOKPLTFNDL 360
QY 361 VKPVCLPNPGMMLQPEOLCWIWSHGATEBKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
Db 361 VKPVCLPNPGMMLQPEOLCWIWSHGATEBKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLQGNVDSQGDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVWF 480
Db 421 TPAMICAGFLQGNVDSQGDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVWF 480
QY 481 TDWIYRQMRADG 492
Db 481 TDWIYRQMRADG 492

RESULT 10

US-10-294-025-932
; Sequence 932, Application US/10294025
; Publication No. US20030185830A1

GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C29
; CURRENT APPLICATION NUMBER: US/10/294,025
; CURRENT FILING DATE: 2002-11-12
; NUMBER OF SEQ ID NOS: 1038
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-294-025-932

Query Match 99.7%; Score 2709; DB 14; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.6e-229;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVVEVHPAQYYPSPVQYAPRVLTOA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVVEVHPAQYYPSPVQYAPRVLTOA 60
QY 61 SNPVVCTOPKSPGTVCTSKTKALCITLTGTLVGAALAAAGLLWKPMGSKCSNSGIEC 120
Db 61 SNPVVCTOPKSPGTVCTSKTKALCITLTGTLVGAALAAAGLLWKPMGSKCSNSGIEC 120
QY 121 DSSGTCINPSNWCDCGVSHCPGDEENRCVRLYGNSFNLQVYSSQKSWHPVCQDDWNNY 180
Db 121 DSSGTCINPSNWCDCGVSHCPGDEENRCVRLYGNSFNLQVYSSQKSWHPVCQDDWNNY 180
QY 181 GRAACRDGMGYNKXNFYSSQGIYDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDGMGYNKXNFYSSQGIYDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLSSRSQRIYVGGESALPGAWPQVSLHVQNWHVCGGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLSSRSQRIYVGGESALPGAWPQVSLHVQNWHVCGGSIITPEWIVTAAHCVEK 300
QY 301 PLNNPWHMTAFAGILRQSFYFAGYQVEKVI SHPNYDSKTKNNDIAMKLOKPLTFNDL 360
Db 301 PLNNPWHMTAFAGILRQSFYFAGYQVEKVI SHPNYDSKTKNNDIAMKLOKPLTFNDL 360
QY 361 VKPVCLPNPGMMLQPEOLCWIWSHGATEBKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
Db 361 VKPVCLPNPGMMLQPEOLCWIWSHGATEBKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLQGNVDSQGDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVWF 480

Db 421 TPAMICAGFLQGNVDSQCGSGGLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNVWF 480
QY 481 TDWIYRQMRADG 492
Db 481 TDWIYRQMRADG 492

RESULT 11

US-10-334-038-6
; Sequence 6, Application US/10334038
; Publication No. US20030138865A1
; GENERAL INFORMATION:
; APPLICANT: Macbeth, Kyle J.
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: NUCLEIC ACIDS AND POLYPEPTIDES USEFUL IN
; FILE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF PROSTATE CANCER
; FILE REFERENCE: 07334-135001
; CURRENT APPLICATION NUMBER: US/10/334,038
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: US/09/408,628A
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: US 09/164,159
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: US 09/163,759
; PRIOR FILING DATE: 1998-09-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-334-038-6

Query Match 99.5%; Score 2704; DB 14; Length 492;
Best Local Similarity 99.4%; Pred. No. 7.1e-229;
Matches 489; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALNSGSPALGPYYENHGYQENPYPAQPTVPTVEVHPAQYPPVPOYAPRVLTOA 60
Db 1 MALNSGSPALGPYYENHGYQENPYPAQPTVPTVEVHPAQYPPVPOYAPRVLTOA 60
QY 61 SNPVVCTQPKSPSGTCTCTKTKALCITLTGLTFLVGAALAAAGLLWFKMGSKCSNSGIEC 120
Db 61 SNPVVCTQPKSPSGTCTCTKTKALCITLTGLTFLVGAALAAAGLLWFKMGSKCSNSGIEC 120
QY 121 DSSGTCINPSNWCDCGVSHCPGGEDENCVRLYGNFVLQVYSSQKSWHPVCCDDWNNY 180
Db 121 DSSGTCINPSNWCDCGVSHCPGGEDENCVRLYGNFVLQVYSSQKSWHPVCCDDWNNY 180
QY 181 GRAACRDMGYKNFYSSQGI VDDSGSTSPFKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNFYSSQGI VDDSGSTSPFKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSSRQSRIVGGESALPGAWPMQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLNSSRQSRIVGGESALPGAWPMQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
QY 301 PLNPNPWHWTAFAGILRQSFMEYAGYQVQKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
Db 301 PLNPNPWHWTAFAGILRQSFMEYAGYQVQKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
QY 361 VKPVCILPNPQMMLOPEQLCWISGWGATEEKGKTSSEVLNAKVLIIETQRCNSRYVYDNL 420
Db 361 VKPVCILPNPQMMLOPEQLCWISGWGATEEKGKTSSEVLNAKVLIIETQRCNSRYVYDNL 420
QY 421 TPAMICAGFLQGNVDSQCGSGGLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNVWF 480
Db 421 TPAMICAGFLQGNVDSQCGSGGLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNVWF 480
QY 481 TDWIYRQMRADG 492
Db 481 TDWIYRQMRADG 492

RESULT 12

US-09-825-751A-63
; Sequence 63, Application US/09825751A
; Publication No. US20030065140A1
; GENERAL INFORMATION:
; APPLICANT: Curagen Corporation
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Fernandes, Elma R.
; APPLICANT: Taupier, Raymond J.
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Spyttek, Kimberly A.
; APPLICANT: Rastelli, Luca
; APPLICANT: Herltan, John L.
; TITLE OF INVENTION: Novel Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-750
; CURRENT APPLICATION NUMBER: US/09/825,751A
; CURRENT FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/194,314
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/225,693
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-825-751A-63

Query Match 99.4%; Score 2701; DB 12; Length 492;
Best Local Similarity 99.0%; Pred. No. 1.3e-228;
Matches 487; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALNSGSPALGPYYENHGYQENPYPAQPTVPTVEVHPAQYPPVPOYAPRVLTOA 60
Db 1 MALNSGSPALGPYYENHGYQENPYPAQPTVPTVEVHPAQYPPVPOYAPRVLTOA 60
QY 61 SNPVVCTQPKSPSGTCTCTKTKALCITLTGLTFLVGAALAAAGLLWFKMGSKCSNSGIEC 120
Db 61 SNPVVCTQPKSPSGTCTCTKTKALCITLTGLTFLVGAALAAAGLLWFKMGSKCSNSGIEC 120
QY 121 DSSGTCINPSNWCDCGVSHCPGGEDENCVRLYGNFVLQVYSSQKSWHPVCCDDWNNY 180
Db 121 DSSGTCINPSNWCDCGVSHCPGGEDENCVRLYGNFVLQVYSSQKSWHPVCCDDWNNY 180
QY 181 GRAACRDMGYKNFYSSQGI VDDSGSTSPFKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNFYSSQGI VDDSGSTSPFKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSSRQSRIVGGESALPGAWPMQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLNSSRQSRIVGGESALPGAWPMQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
QY 301 PLNPNPWHWTAFAGILRQSFMEYAGYQVQKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
Db 301 PLNPNPWHWTAFAGILRQSFMEYAGYQVQKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
QY 361 VKPVCILPNPQMMLOPEQLCWISGWGATEEKGKTSSEVLNAKVLIIETQRCNSRYVYDNL 420
Db 361 VKPVCILPNPQMMLOPEQLCWISGWGATEEKGKTSSEVLNAKVLIIETQRCNSRYVYDNL 420
QY 421 TPAMICAGFLQGNVDSQCGSGGLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNVWF 480
Db 421 TPAMICAGFLQGNVDSQCGSGGLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNVWF 480
QY 481 TDWIYRQMRADG 492
Db 481 TDWIYRQMRADG 492

RESULT 13

US-09-759-143-895
; Sequence 895, Application US/09759143
; Patent No. US200202248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-143-895

Query Match 99.2%; Score 2696; DB 9; Length 492;
Best Local Similarity 98.8%; Pred. No. 3.6e-228;
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVYVHVPAYQYPSVPQYAPRVLTQA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVYVHVPAYQYPSVPQYAPRVLTQA 60
QY 61 SNPVCTQPKSPSTGCTCTKALCITTLGTFLVGAALAAAGLLKFKGSKCSNGIEC 120
Db 61 SNPVCTQPKSPSTGCTCTKALCITTLGTFLVGAALAAAGLLKFKGSKCSNGIEC 120
QY 121 DSSGTCINPSNWCDSVSHCPGDENRCVRLYGPNFILQVYSSQRKSWHPVCDDWNNY 180
Db 121 DSSGTCINPSNWCDSVSHCPGDENRCVRLYGPNFILQVYSSQRKSWHPVCDDWNNY 180
QY 181 GRAACRDMGYKNFYSSQGI VDDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAVVSLR 240
Db 181 GRAACRDMGYKNFYSSQGI VDDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
QY 301 PLNNPWHWTAFAGILRQSFMYGAGYQVKEVISHPNYDSKTKNDIALMKLOKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFMYGAGYQVKEVISHPNYDSKTKNDIALMKLOKPLTFNDL 360
QY 361 VKPVCPLNPGMMLQPEQLCWSGWGATEKGTSEVLNAAKVLLIETQCNRYVVDNLI 420
Db 361 VKPVCPLNPGMMLQPEQLCWSGWGATEKGTSEVLNAAKVLLIETQCNRYVVDNLI 420
QY 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWLLIGDTSWGGCAKAYRPGYGVNVMVF 480
Db 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWLLIGDTSWGGCAKAYRPGYGVNVMVF 480
QY 481 TDWIYRQWADG 492
Db 481 TDWIYRQWKANG 492

RESULT 14

US-09-780-669-895
; Sequence 895, Application US/09780669
; Patent No. US20020051977A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C24
; CURRENT APPLICATION NUMBER: US/09/780,669
; CURRENT FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 943
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-780-669-895

Query Match 99.2%; Score 2696; DB 9; Length 492;
Best Local Similarity 98.8%; Pred. No. 3.6e-228;
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVYVHVPAYQYPSVPQYAPRVLTQA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVYVHVPAYQYPSVPQYAPRVLTQA 60
QY 61 SNPVCTQPKSPSTGCTCTKALCITTLGTFLVGAALAAAGLLKFKGSKCSNGIEC 120
Db 61 SNPVCTQPKSPSTGCTCTKALCITTLGTFLVGAALAAAGLLKFKGSKCSNGIEC 120
QY 121 DSSGTCINPSNWCDSVSHCPGDENRCVRLYGPNFILQVYSSQRKSWHPVCDDWNNY 180
Db 121 DSSGTCINPSNWCDSVSHCPGDENRCVRLYGPNFILQVYSSQRKSWHPVCDDWNNY 180
QY 181 GRAACRDMGYKNFYSSQGI VDDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAVVSLR 240
Db 181 GRAACRDMGYKNFYSSQGI VDDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
QY 301 PLNNPWHWTAFAGILRQSFMYGAGYQVKEVISHPNYDSKTKNDIALMKLOKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFMYGAGYQVKEVISHPNYDSKTKNDIALMKLOKPLTFNDL 360
QY 361 VKPVCPLNPGMMLQPEQLCWSGWGATEKGTSEVLNAAKVLLIETQCNRYVVDNLI 420
Db 361 VKPVCPLNPGMMLQPEQLCWSGWGATEKGTSEVLNAAKVLLIETQCNRYVVDNLI 420
QY 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWLLIGDTSWGGCAKAYRPGYGVNVMVF 480

Db 481 TDWIYROMKANG 492

Search completed: June 1, 2004, 14:48:07
Job time : 125 secs

Db 421 TPAMICAGFLQGNVDSQDGGSGPLVTSNNNIWLIIGDTSWGSCKAKAYRPGVYGNVMVF 480

Qy 481 TDWIYROMKANG 492

Db 481 TDWIYROMKANG 492

RESULT 15

US-09-879-792-14

/ Sequence 14, Application US/09879792

/ Patent No. US20020061850A1

/ GENERAL INFORMATION:

/ APPLICANT: Xiao, Yonghong

/ APPLICANT: Gedrich, Richard

/ TITLE OF INVENTION: Regulation of Human Transmembrane Serine

/ TITLE OF INVENTION: Protease

/ FILE REFERENCE: 02973.00035

/ CURRENT APPLICATION NUMBER: US/09/879,792

/ CURRENT FILING DATE: 2001-06-13

/ PRIOR APPLICATION NUMBER: US 60/211,224

/ PRIOR FILING DATE: 2000-06-13

/ PRIOR APPLICATION NUMBER: US 60/283,353

/ PRIOR FILING DATE: 2001-04-13

/ PRIOR APPLICATION NUMBER: US 60/283,648

/ PRIOR FILING DATE: 2001-04-16

/ PRIOR APPLICATION NUMBER: PCT

/ PRIOR FILING DATE: 2001-06-12

/ NUMBER OF SEQ ID NOS: 36

/ SOFTWARE: FastSeq for Windows Version 4.0

/ SEQ ID NO 14

/ LENGTH: 492

/ TYPE: PRT

/ ORGANISM: Homo sapiens

US-09-879-792-14

(Docket No. US20020061850A1 LIO-81-WO)

Query Match 99.2%; Score 2696; DB 9; Length 492;
Best Local Similarity 98.8%; Pred. No. 3.6e-228;
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVVHHPAQYYPSPVQYAPRVLTOA 60

Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVVHHPAQYYPSPVQYAPRVLTOA 60

Qy 61 SNPVVCTOPKSPGTVCTSKTKALCITLTGTLVGAALAAAGLLKFMGSKCSNSGIEC 120

Db 61 SNPVVCTOPKSPGTVCTSKTKALCITLTGTLVGAALAAAGLLKFMGSKCSNSGIEC 120

Qy 121 DSSGTCINPSNWCDCVSHCPGSEDENRCVLYGNPFIQVYSSORKSWHPVCQDDWNEY 180

Db 121 DSSGTCINPSNWCDCVSHCPGSEDENRCVLYGNPFIQVYSSORKSWHPVCQDDWNEY 180

Qy 181 GRAACRDMGYKKNFYSSQGIYVDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAVVSLR 240

Db 181 GRAACRDMGYKKNFYSSQGIYVDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAVVSLR 240

Qy 241 CIACGVNLNSSRQSRIVGSEALPGAWPQVSLHVQNVHVCVGGSIITPEWITVTAACHVEK 300

Db 241 CIACGVNLNSSRQSRIVGSEALPGAWPQVSLHVQNVHVCVGGSIITPEWITVTAACHVEK 300

Qy 301 PLNNPWHMTAFAGILRQSPMFYAGYQVEKVIHSHENYDSKYNNDIALMKLQKPLTFNDL 360

Db 301 PLNNPWHMTAFAGILRQSPMFYAGYQVEKVIHSHENYDSKYNNDIALMKLQKPLTFNDL 360

Qy 361 VKPVCLPNPQWMLQBPQLCWIISGWGATEBKGTSEVLNAKVLLIETQRCNSRYVDNLI 420

Db 361 VKPVCLPNPQWMLQBPQLCWIISGWGATEBKGTSEVLNAKVLLIETQRCNSRYVDNLI 420

Qy 421 TPAMICAGFLQGNVDSQDGGSGPLVTSNNNIWLIIGDTSWGSCKAKAYRPGVYGNVMVF 480

Db 421 TPAMICAGFLQGNVDSQDGGSGPLVTSNNNIWLIIGDTSWGSCKAKAYRPGVYGNVMVF 480

Qy 481 TDWIYROMKANG 492

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 14:04:23 ; Search time 136 Seconds
(without alignments)
1022.157 Million cell updates/sec

Title: US-09-615-285B-2

Perfect score: 217
Sequence: 1 MALNSGSPPAIGPYENHGY.....VIGNVMVFTDIYQMRADG 492

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04 : *
1: geneseqp1980s : *
2: geneseqp1990s : *
3: geneseqp2000s : *
4: geneseqp2001s : *
5: geneseqp2002s : *
6: geneseqp2003as : *
7: geneseqp2003bs : *
8: geneseqp2004s : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2717	100.0	492	AA44406	AA44406 Human 20P
2	2717	100.0	492	AA77726	AA77726 Human tum
3	2717	100.0	492	AAE18096	AAE18096 Human 20P
4	2714	99.9	492	AAE18098	AAE18098 Human 20P
5	2713	99.9	492	AAE18100	AAE18100 Human 20P
6	2713	99.9	492	AAE18099	AAE18099 Human 20P
7	2713	99.9	492	AAE18101	AAE18101 Human 20P
8	2711	99.8	492	ADB75591	ADB75591 Prostate
9	2709	99.7	492	AA92050	AA92050 HrpCa6/7
10	2709	99.7	492	AA92050	AA92050 HrpCa6/7
11	2709	99.7	492	AA92050	AA92050 HrpCa6/7
12	2709	99.7	492	AA92050	AA92050 HrpCa6/7
13	2709	99.7	492	AA92050	AA92050 HrpCa6/7
14	2709	99.7	492	AA92050	AA92050 HrpCa6/7
15	2709	99.7	492	AA92050	AA92050 HrpCa6/7
16	2696	99.2	492	AA92050	AA92050 HrpCa6/7
17	2696	99.2	492	AA92050	AA92050 HrpCa6/7
18	2696	99.2	492	AA92050	AA92050 HrpCa6/7
19	2696	99.2	492	AA92050	AA92050 HrpCa6/7
20	2696	99.2	492	AA92050	AA92050 HrpCa6/7
21	2696	99.2	492	AA92050	AA92050 HrpCa6/7
22	2696	99.2	492	AA92050	AA92050 HrpCa6/7
23	2696	99.2	492	AA92050	AA92050 HrpCa6/7
24	2696	99.2	492	AA92050	AA92050 HrpCa6/7
25	2696	99.2	492	AA92050	AA92050 HrpCa6/7

26	2696	99.2	492	6	ABR54511	Human tra
27	2696	99.2	492	7	ADB14345	Human tra
28	2171	79.9	393	4	AAW01317	AAW01317 P100C pa
29	2171	79.9	393	4	AAU69962	AAU69962 Human pro
30	2171	79.9	393	4	ABU71853	ABU71853 Prostate
31	2171	79.9	393	5	ABB95422	ABB95422 Human P10
32	2171	79.9	393	5	ABR54534	ABR54534 Prostate
33	2171	79.9	393	7	ADB14384	Human pro
34	2126	78.2	384	6	ADAL0905	ADAL0905 Human CDN
35	1540	56.7	283	3	AA981492	AA981492 Human pro
36	1540	56.7	283	5	ABB78328	ABB78328 Amino aci
37	1165	42.9	209	4	AAW01295	AAW01295 Human tra
38	1165	42.9	209	4	AAU69940	AAU69940 Human tra
39	1165	42.9	209	4	ABU71831	ABU71831 Prostate
40	1165	42.9	209	5	ABR54540	Human tra
41	1165	42.9	209	6	ABR54512	Partial H
42	1165	42.9	209	7	ADB14347	Human tra
43	1165	42.9	209	7	ADB14347	Human tra
44	1019.5	37.5	418	6	ABP56845	Consensus
45	888	32.7	454	3	AB332246	Tumour as

ALIGNMENTS

RESULT 1
AA44406
ID AA44406 standard; protein; 492 AA.
XX AC AA44406;
XX AC AA44406;
XX 22-MAR-2000 (first entry)
XX DE Human 20P1F12-GTC2 protein.
XX 20P1F12; TMPSR2; androgen; serine protease; 20P1F12-GTC1; cancer;
XX transmembrane protein; colon; prostate; prostate tumour.
XX Homo sapiens.
XX WO9962942-A2.
XX 09-DEC-1999.
XX 01-JUN-1999; 99WO-US012253.
XX 01-JUN-1998; 98US-0087598P.
XX 29-JUN-1998; 98US-0091474P.
XX 14-APR-1999; 99US-0129521P.
XX (UROC-) UROGENESYS INC.
XX (AFAR/) AFAR D E.
XX (HUBE/) HUBERT R S.
XX (LEON/) LEONG K.
XX (RAIT/) RAITANO A B.
XX (SAFF/) SAFFRAN D C.
XX Afar DE, Hubert RS, Leong K, Raitano AB, Safran DC;
XX WPI; 2000-116363/10.
XX N-PSDB; AA229636.
XX Novel cell surface antigen useful to treat colon and prostate cancer.
XX Claim 1; Fig 1; 58pp; English.

The present sequence is the 20P1F12 protein (also known as the TMPSR2 protein) which is a prostate-specific, androgen-regulated, cell surface serine protease. It is a glycosylated type II transmembrane protein with an extracellular C-terminal serine protease domain, a scavenger receptor cysteine-rich domain, an LDL receptor class A domain and a predicted transmembrane domain. Host cells can be transformed to produce this protein, using vector containing 20P1F12/TMPSR2 gene (also designated

KW Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTC1; cell growth; neoplasm;
 KW cancer; vaccine; human.
 XX
 OS Homo sapiens.
 XX
 PN WO200204953-A2.
 XX
 PD 17-JAN-2002.
 XX
 PF 12-JUL-2001; 2001WO-US022168.
 XX
 PR 12-JUL-2000; 2000US-00615285.
 XX
 PA (AGEN-) AGENSYS INC.
 XX
 PI Saferan D, Raitano AB, Hubert RS, Jakobovits A, Faris M;
 PI Challita-Eid PM;
 XX
 DR WPI; 2002-154967/20.
 XX
 DR N-PSDB; AAD28778.
 XX
 PT Examining a biological sample for evidence of dysregulated cellular
 PT growth, comprises comparing the status of prostate-specific, androgen-
 PT regulated, secreted serine protease, 20P1F12/TMPRSS2, in a corresponding
 PT normal sample.
 XX
 PS Claim 40; Fig 1; 161pp; English.
 XX
 CC The present invention relates to methods and compositions for the
 CC diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
 CC kidney cancer derived from or based on a normally prostate-specific,
 CC androgen regulated, cell membrane associated secreted serine protease
 CC termed 20P1F12/TMPRSS2. The invention further relates to a method of
 CC examining a biological sample for evidence of dysregulated cellular
 CC growth comprises comparing the status of 20P1F12/TMPRSS2 gene (also
 CC designated 20P1F12-GTC1) in the sample to the status of 20P1F12/TMPRSS2
 CC in a corresponding normal sample. The invention also relates to 20P1F12/
 CC TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
 CC invention are used for examining a sample such as blood, serum, stool,
 CC urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
 CC The dysregulated cell growth is indicative of bladder cancer, lung
 CC cancer, kidney cancer or ovarian cancer. It is useful for identifying
 CC evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
 CC portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
 CC expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
 CC lung cancer, ovarian cancer or metastatic cancer. The present sequence
 CC is human 20P1F12-GTC1 protein
 XX
 SQ Sequence 492 AA;
 Query Match 100.0%; Score 2717; DB 5; Length 492;
 Best Local Similarity 100.0%; Pred. No. 6.7e-203; Mismatches 0; Gaps 0;
 Matches 492; Conservative 0; Indels 0; Gaps 0;
 QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVYEVHPAQYPSVPQYAPRVLTQA 60
 DB 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVYEVHPAQYPSVPQYAPRVLTQA 60
 QY 61 SNPVVCTQPKSPGTCVCTSKTKKALCITLTGTLVGAALAGLLKFMGSKNSGIEC 120
 DB 61 SNPVVCTQPKSPGTCVCTSKTKKALCITLTGTLVGAALAGLLKFMGSKNSGIEC 120
 QY 121 DSSGTCINPNSWCDGVSHCPGGEENRCVRLYGNPFIQVYSSQKSHHPVQDDWNEY 180
 DB 121 DSSGTCINPNSWCDGVSHCPGGEENRCVRLYGNPFIQVYSSQKSHHPVQDDWNEY 180
 QY 181 GRAACRDGMGYKXNFYSSQGIYVDDSGSTSPKMLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
 DB 181 GRAACRDGMGYKXNFYSSQGIYVDDSGSTSPKMLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
 QY 241 CIACGVNLNSRQSRIVGESALPGAWPQVSLHVQNVHVCQGSITPEWTVTAHCVCK 300
 DB 241 CIACGVNLNSRQSRIVGESALPGAWPQVSLHVQNVHVCQGSITPEWTVTAHCVCK 300

QY 301 PLNNPWHWTAFAGILRQSEFMFYGAGYQVEKVI SHPNYDSKTKNNDIALMKLOKPLTFNDL 360
 DB 301 PLNNPWHWTAFAGILRQSEFMFYGAGYQVEKVI SHPNYDSKTKNNDIALMKLOKPLTFNDL 360
 QY 361 VKPVCLPNPGMMLQPEQLCWISGWGATEBKGTSEVLNAKVLLIETQCNRSYVVDNLI 420
 DB 361 VKPVCLPNPGMMLQPEQLCWISGWGATEBKGTSEVLNAKVLLIETQCNRSYVVDNLI 420
 QY 421 TPAMICAGFLQGNVDSQCGDGGPLVTSKNTNWLIGDTSWGGCAKAYRPGVYGNVMVF 480
 DB 421 TPAMICAGFLQGNVDSQCGDGGPLVTSKNTNWLIGDTSWGGCAKAYRPGVYGNVMVF 480
 QY 481 TDWIYRQMRADG 492
 DB 481 TDWIYRQMRADG 492
 RESULT 4
 AAE18098
 ID AAE18098 standard; protein; 492 AA.
 XX
 AC AAE18098;
 XX
 DT 07-MAY-2002 (first entry)
 XX
 XX Human 20P1F12/TMPRSS2 mutant protein (S441A).
 XX
 DE Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTC1; cell growth; neoplasm;
 KW cancer; vaccine; human.
 KW
 OS Homo sapiens.
 OS
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 441
 FT /note= "Wild type Ser substituted with Ala"
 XX
 XX WO200204953-A2.
 XX
 PD 17-JAN-2002.
 XX
 PF 12-JUL-2001; 2001WO-US022168.
 XX
 PR 12-JUL-2000; 2000US-00615285.
 XX
 PA (AGEN-) AGENSYS INC.
 XX
 PI Saferan D, Raitano AB, Hubert RS, Jakobovits A, Faris M;
 PI Challita-Eid PM;
 XX
 DR WPI; 2002-154967/20.
 XX
 PT Examining a biological sample for evidence of dysregulated cellular
 PT growth, comprises comparing the status of prostate-specific, androgen-
 PT regulated, secreted serine protease, 20P1F12/TMPRSS2, in a corresponding
 PT normal sample.
 XX
 PS Example 10; Page; 161pp; English.
 XX
 CC The present invention relates to methods and compositions for the
 CC diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
 CC kidney cancer derived from or based on a normally prostate-specific,
 CC androgen regulated, cell membrane associated secreted serine protease
 CC termed 20P1F12/TMPRSS2. The invention further relates to a method of
 CC examining a biological sample for evidence of dysregulated cellular
 CC growth comprises comparing the status of 20P1F12/TMPRSS2 gene (also
 CC designated 20P1F12-GTC1) in the sample to the status of 20P1F12/TMPRSS2
 CC in a corresponding normal sample. The invention also relates to 20P1F12/
 CC TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
 CC invention are used for examining a sample such as blood, serum, stool,
 CC urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
 CC The dysregulated cell growth is indicative of bladder cancer, lung

CC cancer, kidney cancer or ovarian cancer. It is useful for identifying
CC evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
CC portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
CC expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
CC lung cancer, ovarian cancer or metastatic cancer. The present sequence is
CC human 20P1F12/TMPRSS2 mutant protein (S441A). Note: This sequence is not
CC shown in the specification but is derived from wild type 20P1F12/TMPRSS2
CC protein shown as SEQ ID NO: 2 in figure 1 of the specification (AAE18096)
XX
SQ Sequence 492 AA;

Query Match 99.9%; Score 2714; DB 5; Length 492;
Best Local Similarity 99.8%; Pred. No. 1.4e-202;
Matches 491; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVYEVHPAQYYPSPVQYAPRVLTQA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVYEVHPAQYYPSPVQYAPRVLTQA 60

Qy 61 SNPVVCTQPKSPSGTCTSKTKALCITLTGLTFLVGAALAAAGLLKFMGSKCSNSGIEC 120
Db 61 SNPVVCTQPKSPSGTCTSKTKALCITLTGLTFLVGAALAAAGLLKFMGSKCSNSGIEC 120

Qy 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGNFNLQVYSSQKSWHPVCQDDWNNY 180
Db 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGNFNLQVYSSQKSWHPVCQDDWNNY 180

Qy 181 GRAACRDMGYKKNFYSSQGIYVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKKNFYSSQGIYVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240

Qy 241 CIACGVNLNSRQSRIVGGESALPGAMPQVSLHVQNVHVCVGGSIITPEWIVTAACHVEK 300
Db 241 CIACGVNLNSRQSRIVGGESALPGAMPQVSLHVQNVHVCVGGSIITPEWIVTAACHVEK 300

Qy 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVIHSHPNYDSKTKNQNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVIHSHPNYDSKTKNQNDIALMKLQKPLTFNDL 360

RESULT 5
AAE18100
ID AAE18100 standard; protein; 492 AA.
XX
AC AAE18100;
XX
DT 07-MAY-2002 (first entry)
XX
DE Human 20P1F12/TMPRSS2 mutant protein (R252Q).
XX
KW Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTC1; cell growth; neoplasm;
KW cancer; vaccine; human.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 252
XX /note= "Wild type Arg substituted with Gln"
PN WO200204953-A2.

XX 17-JAN-2002.
XX 12-JUL-2001; 2001WO-US022168.
XX 12-JUL-2000; 2000US-00615285.
XX (AGEN-) AGENSYS INC.
XX
XX Saferan D, Raitano AB, Hubert RS, Jakobovits A, Faris M;
XX Challita-Eid PM;
XX WPI; 2002-154967/20.
XX
XX Examining a biological sample for evidence of dysregulated cellular
XX growth, comprises comparing the status of prostate-specific, androgen-
XX regulated, secreted serine protease, 20P1F12/TMPRSS2, in a corresponding
XX normal sample.
XX
XX Example 10; Page; 161pp; English.
XX
XX The present invention relates to methods and compositions for the
XX diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
XX kidney cancer derived from or based on a normally prostate-specific,
XX androgen regulated, cell membrane associated secreted serine protease
XX termed 20P1F12/TMPRSS2. The invention further relates to a method of
XX examining a biological sample for evidence of dysregulated cellular
XX growth comprises comparing the status of 20P1F12/TMPRSS2 gene (also
XX designated 20P1F12-GTC1) in the sample to the status of 20P1F12/TMPRSS2
XX in a corresponding normal sample. The invention also relates to 20P1F12/
XX TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
XX invention are used for examining a sample such as blood, serum, stool,
XX urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
XX The dysregulated cell growth is indicative of bladder cancer, lung
XX cancer, kidney cancer or ovarian cancer. It is useful for identifying
XX evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
XX portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
XX expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
XX lung cancer, ovarian cancer or metastatic cancer. The present sequence is
XX human 20P1F12/TMPRSS2 mutant protein (R252Q). Note: This sequence is not
XX shown in the specification but is derived from wild type 20P1F12/TMPRSS2
XX protein shown as SEQ ID NO: 2 in figure 1 of the specification (AAE18096)
XX
SQ Sequence 492 AA;

Query Match 99.9%; Score 2713; DB 5; Length 492;
Best Local Similarity 99.8%; Pred. No. 1.4e-202;
Matches 491; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVYEVHPAQYYPSPVQYAPRVLTQA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVYEVHPAQYYPSPVQYAPRVLTQA 60

Qy 61 SNPVVCTQPKSPSGTCTSKTKALCITLTGLTFLVGAALAAAGLLKFMGSKCSNSGIEC 120
Db 61 SNPVVCTQPKSPSGTCTSKTKALCITLTGLTFLVGAALAAAGLLKFMGSKCSNSGIEC 120

Qy 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGNFNLQVYSSQKSWHPVCQDDWNNY 180
Db 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGNFNLQVYSSQKSWHPVCQDDWNNY 180

Qy 181 GRAACRDMGYKKNFYSSQGIYVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKKNFYSSQGIYVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240

Qy 241 CIACGVNLNSRQSRIVGGESALPGAMPQVSLHVQNVHVCVGGSIITPEWIVTAACHVEK 300
Db 241 CIACGVNLNSRQSRIVGGESALPGAMPQVSLHVQNVHVCVGGSIITPEWIVTAACHVEK 300

Qy 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVIHSHPNYDSKTKNQNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVIHSHPNYDSKTKNQNDIALMKLQKPLTFNDL 360

QY 361 VKFVCLPFGMMLOPEQLCWISGWGATEBKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
Db 361 VKFVCLPFGMMLOPEQLCWISGWGATEBKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWMLLIGDTSWGSCKAKAYRPGVYGNMVF 480
Db 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWMLLIGDTSWGSCKAKAYRPGVYGNMVF 480
QY 481 TDWIYQMRADG 492
Db 481 TDWIYQMRADG 492

RESULT 6
AAE18099
ID AAE18099 standard; protein; 492 AA.
XX AAE18099;
XX 07-MAY-2002 (first entry)
XX Human 20P1F12/TMPRSS2 mutant protein (R240Q).
XX Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTCL; cell growth; neoplasm;
XX cancer; vaccine; human.
XX Homo sapiens.
XX Synthetic.
FH Key Location/Qualifiers
FT Misc-difference 240 /note= "Wild type Arg substituted with Gln"
FT 20020204953-A2.
XX WO200204953-A2.
XX 17-JAN-2002.
XX 12-JUL-2001; 2001WO-US022168.
XX 12-JUL-2000; 2000US-00615285.
XX (AGEN-) AGENSYS INC.
XX Saferran D, Raitano AB, Hubert RS, Jakobovits A, Paris M;
XX Challita-Eid PM;
XX WPI; 2002-154967/20.
XX Examining a biological sample for evidence of dysregulated cellular
XX growth, comprises comparing the status of prostate-specific, androgen-
XX regulated, secreted serine protease, 20P1F12/TMPRSS2, in a corresponding
XX normal sample.
XX Example 10; Page; 161pp; English.
XX The present invention relates to methods and compositions for the
XX diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
XX kidney cancer derived from or based on a normally prostate-specific,
XX androgen regulated, cell membrane associated secreted serine protease
XX termed 20P1F12/TMPRSS2. The invention further relates to a method of
XX examining a biological sample for evidence of dysregulated cellular
XX growth comprising comparing the status of 20P1F12/TMPRSS2 gene (also
XX designated 20P1F12-GTCL) in the sample to the status of 20P1F12/TMPRSS2
XX in a corresponding normal sample. The invention also relates to 20P1F12/
XX TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
XX invention are used for examining a sample such as blood, serum, stool,
XX urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
XX The dysregulated cell growth is indicative of bladder cancer, lung
XX cancer, kidney cancer or ovarian cancer. It is useful for identifying
XX evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
XX portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
XX expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
XX lung cancer, ovarian cancer or metastatic cancer. The present sequence is

CC human 20P1F12/TMPRSS2 mutant protein (R240Q). Note: This sequence is not
CC shown in the specification but is derived from wild type 20P1F12/TMPRSS2
CC protein shown as SEQ ID NO: 2 in figure 1 of the specification (AAE18096)
XX
SQ Sequence 492 AA;
Query Match 99.9%; Score 2713; DB 5; Length 492;
Best Local Similarity 99.8%; Pred. No. 1.4e-202; Indels 0; Gaps 0;
Matches 491; Conservative 1; Mismatches 0;
QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVVPTVYEVHQAQYFSPVQYAPRVLTA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVVPTVYEVHQAQYFSPVQYAPRVLTA 60
QY 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAGLLNKFWSKCSNSGIEC 120
Db 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAGLLNKFWSKCSNSGIEC 120
QY 121 DSSGTCTINPSNWCDSVSHCPGGEDENRCVRLYGPENFLOVYSSQKSWHPVCCDDWNEY 180
Db 121 DSSGTCTINPSNWCDSVSHCPGGEDENRCVRLYGPENFLOVYSSQKSWHPVCCDDWNEY 180
QY 181 GRAACRDMGYKNNFYSSQGIIVDSGSTSPMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNNFYSSQGIIVDSGSTSPMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLSSRSQSRIVGSGESALPGAWPQVSLHVQNVHVCVGGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLSSRSQSRIVGSGESALPGAWPQVSLHVQNVHVCVGGSIITPEWIVTAAHCVEK 300
QY 301 PLANNPWHWTAFAGILRQSPFMFYGAGYQVEKVIHSHPNYDSKTKNDIALMKLQKPLTFNDL 360
Db 301 PLANNPWHWTAFAGILRQSPFMFYGAGYQVEKVIHSHPNYDSKTKNDIALMKLQKPLTFNDL 360
QY 361 VKPVCLPFGMMLOPEQLCWISGWGATEBKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
Db 361 VKPVCLPFGMMLOPEQLCWISGWGATEBKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWMLLIGDTSWGSCKAKAYRPGVYGNMVF 480
Db 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWMLLIGDTSWGSCKAKAYRPGVYGNMVF 480
QY 481 TDWIYQMRADG 492
Db 481 TDWIYQMRADG 492

RESULT 7
AAE18101
ID AAE18101 standard; protein; 492 AA.
XX AAE18101;
XX 07-MAY-2002 (first entry)
XX Human 20P1F12/TMPRSS2 mutant protein (R255Q).
XX Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTCL; cell growth; neoplasm;
XX cancer; vaccine; human.
XX Homo sapiens.
XX Synthetic.
XX Key Location/Qualifiers
XX FT Misc-difference 255 /note= "Wild type Arg substituted with Gln"
XX WO200204953-A2.
XX 17-JAN-2002.
XX 12-JUL-2001; 2001WO-US022168.
XX

PR 12-JUL-2000; 2000US-00615285.
XX (AGEN-) AGENSYS INC.
XX Saferran D, Raitano AB, Hubert RS, Jakobovits A, Faris M;
XX Challita-Eid PM;
XX WPI; 2002-154967/20.
XX Examining a biological sample for evidence of dysregulated cellular
XX growth, comprises comparing the status of prostate-specific, androgen-
XX regulated, secreted serine protease, 20P1F12/TPRSS2, in a corresponding
XX normal sample.
XX Example 10; Page; 161pp; English.
XX The present invention relates to methods and compositions for the
XX diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
XX kidney cancer derived from or based on a normally prostate-specific,
XX androgen regulated, cell membrane associated secreted serine protease
XX termed 20P1F12/TPRSS2. The invention further relates to a method of
XX examining a biological sample for evidence of dysregulated cellular
XX growth comprises comparing the status of 20P1F12/TPRSS2 gene (also
XX designated 20P1F12-GRCl) in the sample to the status of 20P1F12/TPRSS2
XX in a corresponding normal sample. The invention also relates to 20P1F12/
XX TPRS2 polynucleotides and their corresponding proteins. Methods of the
XX invention are used for examining a sample such as blood, serum, stool,
XX urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
XX The dysregulated cell growth is indicative of bladder cancer, lung
XX cancer, kidney cancer or ovarian cancer. It is useful for identifying
XX evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
XX portion of 20P1F12/TPRSS2 are useful for inhibiting growth of a cell
XX expressing 20P1F12/TPRSS2 in a patient suffering from bladder cancer,
XX lung cancer, ovarian cancer or metastatic cancer. The present sequence is
XX human 20P1F12/TPRSS2 mutant protein (R2550). Note: This sequence is not
XX shown in the specification but is derived from wild type 20P1F12/TPRSS2
XX protein shown as SEQ ID NO: 2 in figure 1 of the specification (AAE18096)
XX SQ Sequence 492 AA;
Query Match 99.9%; Score 2713; DB 5; Length 492;
Best Local Similarity 99.8%; Pred. No. 1.4e-202;
Matches 491; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVVEHPAQYPSVPQYAPRVLTA 60
DB 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVVEHPAQYPSVPQYAPRVLTA 60
QY 61 SNPVVCTOPKPSGTVCTSKKALCITLTGLTGLVGAALAGLLWFMGSKCSNSGIEC 120
DB 61 SNPVVCTOPKPSGTVCTSKKALCITLTGLTGLVGAALAGLLWFMGSKCSNSGIEC 120
QY 121 DSSGTCINPSNWCDSVHCPGGEDENRCVRLYGPNFLOVYSSQKSWHPVCQDDWNEY 180
DB 121 DSSGTCINPSNWCDSVHCPGGEDENRCVRLYGPNFLOVYSSQKSWHPVCQDDWNEY 180
QY 191 GRAACRDGYKNFVSSGIVDDSGSTSFMLNTSAGNVDYKLYHSDACSSKAVVSLR 240
DB 191 GRAACRDGYKNFVSSGIVDDSGSTSFMLNTSAGNVDYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGSGSALPGAWPQVSLHVQNVHVCVGGSIITPEWITVAACVCK 300
DB 241 CIACGVNLNSRQSRIVGSGSALPGAWPQVSLHVQNVHVCVGGSIITPEWITVAACVCK 300
QY 301 PLANNPWHWTAFAGILRQSFMFAGYQYEVKVIHSHPNYDSKTKNDIALMKLQKLTNDL 360
DB 301 PLANNPWHWTAFAGILRQSFMFAGYQYEVKVIHSHPNYDSKTKNDIALMKLQKLTNDL 360
QY 361 VKPVCCLPNFGMLOPEQLCWISGCGATEEKGKTSVLNAKVLIIETQRCNSRYVVDNLI 420
DB 361 VKPVCCLPNFGMLOPEQLCWISGCGATEEKGKTSVLNAKVLIIETQRCNSRYVVDNLI 420
QY 421 TPAMTCAGFLQGNVDSGCGPLVTSKNNIWLIGTWSGSCGCAKAYRPGVYGNVWF 480

DB 421 TPAMTCAGFLQGNVDSGCGPLVTSKNNIWLIGTWSGSCGCAKAYRPGVYGNVWF 480
QY 481 TDWIYRQWRADG 492
DB 481 TDWIYRQWRADG 492
RESULT 8
ADB75591
ID ADB75591 standard; protein; 492 AA.
XX ADB75591;
AC ADB75591;
DT 04-DEC-2003 (first entry)
XX Prostate cancer marker protein.
DE Prostate; cancer; cytostatic; gene therapy; marker.
KW Homo sapiens.
OS Homo sapiens.
XX WO2003009814-A2.
XX 06-FEB-2003.
XX 25-JUL-2002; 2002WO-US023913.
XX 25-JUL-2001; 2001US-0307982P.
XX 22-AUG-2001; 2001US-0314356P.
XX 25-SEP-2001; 2001US-0325020P.
XX 12-DEC-2001; 2001US-0341748P.
XX 05-MAR-2002; 2002US-0362158P.
XX (MILL-) MILLENNIUM PHARM INC.
XX Schlegel R, Monahan JE, Endege WO, Gamavarapu M, Gorbacheva B;
XX Hoersch S, Kamatkar S, Woney AM, Glatt K, Zhao X, Anderson D;
XX WPI; 2003-248033/24.
XX New nucleic acid molecule, useful for diagnosing or treating prostate
XX cancer.
XX Disclosure; SEQ ID NO 415; 99pp; English.
XX The invention relates to newly discovered cancer markers associated with
XX the cancerous state of prostate cells. Also disclosed is a method of
XX assessing whether a patient is afflicted with prostate cancer. The method
XX of the invention involves assessing whether a patient is afflicted with
XX prostate cancer by comparing the level of expression of a marker in a
XX patient sample and the normal level of expression of the marker in a
XX control non-prostate cancer sample, where a significant increase in the
XX level of expression of the marker in the patient sample and the normal
XX level indicates that the patient is afflicted with prostate cancer.
XX Nucleic acids of the invention are useful for diagnosing or treating
XX prostate cancer, and may be useful in gene therapy. Sequences given in
XX ADB75177-ADB75631 represent marker cDNA and proteins. Note: The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 492 AA;
Query Match 99.8%; Score 2711; DB 7; Length 492;
Best Local Similarity 99.6%; Pred. No. 2e-202;
Matches 490; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVVEHPAQYPSVPQYAPRVLTA 60
DB 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVVEHPAQYPSVPQYAPRVLTA 60
QY 61 SNPVVCTOPKPSGTVCTSKKALCITLTGLTGLVGAALAGLLWFMGSKCSNSGIEC 120

Db 61 SNPVCTQPKSPGTVCTSKKALCITLTLGTLFLVGAALAGLLWKFWSKCSNSGIEC 120
 Qy 121 DSSGTCINPNSWCDGVSHCPGEGEDENRCVLYGPNFILOVYSSQKSWHPVQDDWNNY 180
 Db 121 DSSGTCINPNSWCDGVSHCPGEGEDENRCVLYGPNFILOVYSSQKSWHPVQDDWNNY 180
 Qy 181 GRAACRDMGYKNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVSLR 240
 Db 181 GRAACRDMGYKNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVSLR 240
 Qy 241 CIACGVNLSSRSQRIYGGESALPGAWPMQVSLHVQNVHVCSSIIITPEWIVTAACHVEK 300
 Db 241 CIACGVNLSSRSQRIYGGESALPGAWPMQVSLHVQNVHVCSSIIITPEWIVTAACHVEK 300
 Qy 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
 Db 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
 Qy 361 VKPVLCPNPGMWLQPEOLCWSGWGATEEKGKTSVLNAKVLIIETQCNRSRYVDNLI 420
 Db 361 VKPVLCPNPGMWLQPEOLCWSGWGATEEKGKTSVLNAKVLIIETQCNRSRYVDNLI 420
 Qy 421 TPAMICAGFLQGNVDSQCGSDGGLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVMVF 480
 Db 421 TPAMICAGFLQGNVDSQCGSDGGLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVMVF 480
 Qy 481 TDWIYRQMRADG 492
 Db 481 TDWIYRQMRADG 492

RESULT 9

AAY92050
 ID AAY92050 standard; protein; 492 AA.

XX AC AAY92050;

XX DT 01-AUG-2000 (first entry)

XX DE HxPCa6/7 polypeptide from androgen-inducible gene clone.

XX KW Androgen inducible; testosterone; prostate cancer; cytostatic; TMPRSS2; diagnosis.

XX OS Homo sapiens.

XX PN WO200018961-A2.

XX PD 06-APR-2000.

XX PF 30-SEP-1999; 99WO-US022535.

XX PR 30-SEP-1998; 98US-00163759.

XX PR 30-SEP-1998; 98US-00164159.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Macbeth KJ, Shyjan AW;

XX DR WFI; 2000-293182/25.

XX DR N-PSDB; AAA08803.

XX PT Novel methods for identifying compounds for treating prostate cancer comprising measuring the level of expression or activity of 1 or more of

XX PT 11 genes or their products.

XX PS Claim 2; Fig 3; 108pp; English.

XX CC This protein is encoded by a gene which is androgen (e.g. testosterone) inducible in androgen-dependent prostate cancer cells (e.g. LNCap cells) and constitutively expressed in androgen-independent prostate cancer cells (e.g. LNCap cells). Agents which decrease the expression or

CC activity of these clones may slow or arrest the growth of prostate cancer cells or may kill them. HxPCa6/7 can be obtained from the sequence of the known gene for TMPRSS2. A compound useful for treating prostate cancer can be identified in a novel method comprising measuring the expression level, or activity, of HxPCa2, 3, 6/7, 8, 9, 10, 13, 14, 15, 19, or peripheral-type benzodiazepine receptor (pBR) in a cell, in the presence and absence of a test compound. The sequences may also be used in diagnosis of prostate cancer and to determine efficacy of treatment for prostate cancer

XX SQ Sequence 492 AA;

Query Match 99.7%; Score 2709; DB 3; Length 492;
 Best Local Similarity 99.8%; Pred. No. 2.8e-202;
 Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MALNSSGSPAIGPYENHGYQENPYPAQPTVTVYEVHPAQYFSPVQVAPVLTQA 60

Db 1 MALNSSGSPAIGPYENHGYQENPYPAQPTVTVYEVHPAQYFSPVQVAPVLTQA 60

Qy 61 SNPVCTQPKSPGTVCTSKKALCITLTLGTLFLVGAALAGLLWKFWSKCSNSGIEC 120

Db 61 SNPVCTQPKSPGTVCTSKKALCITLTLGTLFLVGAALAGLLWKFWSKCSNSGIEC 120

Qy 121 DSSGTCINPNSWCDGVSHCPGEGEDENRCVLYGPNFILOVYSSQKSWHPVQDDWNNY 180

Db 121 DSSGTCINPNSWCDGVSHCPGEGEDENRCVLYGPNFILOVYSSQKSWHPVQDDWNNY 180

Qy 181 GRAACRDMGYKNFYSSQGIIVDDSGSTSPMKLNTSAGNVDIYKLYHSDACSSKAVSLR 240

Db 181 GRAACRDMGYKNFYSSQGIIVDDSGSTSPMKLNTSAGNVDIYKLYHSDACSSKAVSLR 240

Qy 241 CIACGVNLSSRSQRIYGGESALPGAWPMQVSLHVQNVHVCSSIIITPEWIVTAACHVEK 300

Db 241 CIACGVNLSSRSQRIYGGESALPGAWPMQVSLHVQNVHVCSSIIITPEWIVTAACHVEK 300

Qy 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360

Db 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360

Qy 361 VKPVLCPNPGMWLQPEOLCWSGWGATEEKGKTSVLNAKVLIIETQCNRSRYVDNLI 420

Db 361 VKPVLCPNPGMWLQPEOLCWSGWGATEEKGKTSVLNAKVLIIETQCNRSRYVDNLI 420

Qy 421 TPAMICAGFLQGNVDSQCGSDGGLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVMVF 480

Db 421 TPAMICAGFLQGNVDSQCGSDGGLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVMVF 480

Qy 481 TDWIYRQMRADG 492

Db 481 TDWIYRQMRADG 492

RESULT 10

AAM01315
 ID AAM01315 standard; protein; 492 AA.

XX AC AAM01315;

XX DT 04-OCT-2001 (first entry)

XX DE P1000C amino acid sequence.

XX KW Human; prostate cancer; prostate-specific; diagnosis; vaccine; cytostatic; gene therapy; metastasis.

XX OS Homo sapiens.

XX PN WO2000151633-A2.

XX PD 13-JUL-2001.

XX PF 16-JAN-2001; 2001WO-US0001574.

XX 14-JAN-2000; 2000US-00483672.
XX (CORI-) CORIXA CORP.
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
XX Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAM;
XX Wang A, Meagher MJ;
XX WPI; 2001-425873/45.
XX New polynucleotide encoding a prostate-specific protein, for diagnosing,
XX monitoring and treating prostate cancer in a patient and for use in
XX vaccines.
XX Claim 2; Page 539-540; 543pp; English.
XX The present invention describes polynucleotide sequences (I) which encode
XX prostate-specific proteins (II). (I) and (II) have cytostatic activity,
XX and can be used in vaccine production and gene therapy. (I), (II),
XX antibodies to (II), fusion proteins comprising (II), and isolated T cells
XX prepared using (I) or (II) are used treat cancer in a patient. (I) and
XX the antibodies are also used in the detection of cancer in a patient. The
XX cancer that is diagnosed or treated is particularly prostate cancer. (I)
XX and (II) can be used in vaccines. The antibodies or (I) can be used for
XX monitoring the progression of cancer in a patient. (I) and (II) can also
XX be used to improve diagnostic and therapeutic methods for prostate
XX cancer. They can indicate the level of metastasis as well as the prostate
XX volume. AAH93357 to AAH93344 and AAH01115 to AAH01318 represent
XX polynucleotide and amino acid sequences used in the exemplification of
XX the present invention
XX Sequence 492 AA;

Query Match 99.7%; Score 2709; DB 4; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.8e-202;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVVEHVAQYVSPVQYAPRVLTQA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVVEHVAQYVSPVQYAPRVLTQA 60
Qy 61 SNPVCVCTQKPSGTVCTSKTKKALCITLTGTLVGAALAAAGLLKFMGSKCSNGIEC 120
Db 61 SNPVCVCTQKPSGTVCTSKTKKALCITLTGTLVGAALAAAGLLKFMGSKCSNGIEC 120
Qy 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGPNFLOVYSSQKSWHPVCQDDWNEY 180
Db 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGPNFLOVYSSQKSWHPVCQDDWNEY 180
Qy 181 GRAACRDMGYKNFYSSQGIYVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNFYSSQGIYVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Qy 241 CIACGVNLNSRQRIYSGESALPGAWFQVSLHVQNVHVCVGSIIIPENIVTAACHVEK 300
Db 241 CIACGVNLNSRQRIYSGESALPGAWFQVSLHVQNVHVCVGSIIIPENIVTAACHVEK 300
Qy 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVIHSPNDVSKTKNDIAIMKLOKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVIHSPNDVSKTKNDIAIMKLOKPLTFNDL 360
Qy 361 VKPVCILPNPQMWLOPEQLCWISGWGATEBEKTSVLNAAKVLIIETQRCSRYVYDNL 420
Db 361 VKPVCILPNPQMWLOPEQLCWISGWGATEBEKTSVLNAAKVLIIETQRCSRYVYDNL 420
Qy 421 TPAMICAGFLOQNVDSQCGSGGLPLVTSKNNIWLIGDTSWGSGCAKAYRPGVYGNVWF 480
Db 421 TPAMICAGFLOQNVDSQCGSGGLPLVTSKNNIWLIGDTSWGSGCAKAYRPGVYGNVWF 480
Qy 481 TDWIYQNRADG 492
Db 481 TDWIYQNRADG 492

RESULT 11
AAU69960
ID AAU69960 standard; protein; 492 AA.
XX
XX AC AAU69960;
XX 30-JAN-2002 (first entry)
XX
XX Human prostate cDNA encoded protein #86.
XX Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
XX Homo sapiens.
XX WO200173032-A2.
XX
XX 04-OCT-2001.
XX
XX 27-MAR-2001; 2001WO-US009919.
XX
XX 27-MAR-2000; 2000US-00536857.
XX 09-MAY-2000; 2000US-00568100.
XX 12-MAY-2000; 2000US-00570737.
XX 13-JUN-2000; 2000US-00593793.
XX 27-JUN-2000; 2000US-00605783.
XX 09-AUG-2000; 2000US-00636215.
XX 29-AUG-2000; 2000US-00651236.
XX 06-SEP-2000; 2000US-00657279.
XX 02-OCT-2000; 2000US-00679426.
XX 10-OCT-2000; 2000US-00685156.
XX 09-NOV-2000; 2000US-00705729.
XX
XX (CORI-) CORIXA CORP.
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
XX Fanger GR, Retter MW, Stolk JA, Day CH, Vedwick TS, Carter D;
XX Li SX, Wang A, Skeiky YAM, Hepler WT, Henderson RA;
XX
XX WPI; 2001-639232/73.
XX N-PSDB; AAS64178, AAS64179, AAS64180.
XX New human prostate-specific polypeptides and polynucleotides useful for
XX the diagnosis and treatment of cancer, especially prostate cancer.
XX Claim 2; Page 573-574; 579pp; English.
XX The invention relates to isolated prostate-specific polynucleotides,
XX polypeptides, fusion proteins of the polypeptides, antibodies raised
XX against the polypeptides (or antigenic epitopes derived from them) and
XX antigen-presenting cells expressing the polypeptides. The antibodies are
XX useful for detecting the presence of cancer, especially prostate cancer.
XX The polypeptides, polynucleotides and the antigen-presenting cells are
XX useful for stimulating and/or expanding T cells specific for a tumour
XX protein, and for inhibiting the development of cancer especially prostate
XX cancer. Compositions comprising the polynucleotide and/or polypeptide are
XX useful for stimulating an immune response, and for treating cancer. The
XX oligonucleotide is useful for detecting cancer. The present sequence is a
XX prostate specific polypeptide of the invention
XX
XX Sequence 492 AA;

Query Match 99.7%; Score 2709; DB 4; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.8e-202;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVVEHVAQYVSPVQYAPRVLTQA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVVEHVAQYVSPVQYAPRVLTQA 60
Qy 61 SNPVCVCTQKPSGTVCTSKTKKALCITLTGTLVGAALAAAGLLKFMGSKCSNGIEC 120

Db 61 SNPVCTQKPSGTVCTSKTKKALCITLTGLTFLVGAALAAAGLLWFKMGSKCSNSGIEC 120
Qy 121 DSSGTCINPNCWCDGVSCHPCGGEDENRCVRLYGNFILLQVYSSQKSWHPVCOQDWNENY 180
Db 121 DSSGTCINPNCWCDGVSCHPCGGEDENRCVRLYGNFILLQVYSSQKSWHPVCOQDWNENY 180
Qy 181 GRAACRDMGYKNNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKXLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKXLYHSDACSSKAVVSLR 240
Qy 241 CIACGVNLNSRSQSRIVGGESALPGAWPQVSLHVQNVHVCSSGIITPEWIVTAAHCVCK 300
Db 241 CIACGVNLNSRSQSRIVGGESALPGAWPQVSLHVQNVHVCSSGIITPEWIVTAAHCVCK 300
Qy 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVISHPNVDSTKKNNDIALMKLQKELTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVISHPNVDSTKKNNDIALMKLQKELTFNDL 360
Qy 361 VKPVLCPNPGMWLOPEQLCWISGWSGATEBEKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
Db 361 VKPVLCPNPGMWLOPEQLCWISGWSGATEBEKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
Qy 421 TPAMICAGFLQGNVDSQCGSDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVWF 480
Db 421 TPAMICAGFLQGNVDSQCGSDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVWF 480
Qy 481 TDWIYRQMRADG 492
Db 481 TDWIYRQMRADG 492

RESULT 12

ASU71851
ID ABU71851 standard; protein; 492 AA.

XX AC ABU71851;

XX DT 10-JUN-2003 (first entry)

XX DE Prostate cancer associated protein #67.

XX KW Prostate cancer; vaccine; gene therapy; cytostatic; fusion protein;
XX immunogen; cancer; prostate specific antigen; PSA;
XX KW prostatic acid phosphatase; PAP; prostate specific membrane antigen;
XX PSMA.

XX OS Homo sapiens.

XX PN US2002192763-A1.

XX PD 19-DEC-2002.

XX PF 29-JUN-2001; 2001US-00895793.

XX PR 04-OCT-1999; 99US-01574557.

XX PR 04-OCT-2000; 2000US-00679272.

XX PR 28-MAR-2001; 2001US-00822827.

XX XU (XUJ/) XU J.

PA (DILL/) DILLON D C.

PA (MITC/) MITCHAM J L.

PA (HARL/) HARLOCKER S L.

PA (JIAN/) JIANG Y.

PA (KALO/) KALOS M D.

PA (FANG/) FANGER G R.

PA (RETT/) RETTER M W.

PA (STOL/) STOLK J A.

PA (DAYC/) DAY C H.

PA (VEDV/) VEDVICK T S.

PA (CART/) CARTER D.

PA (LISX/) LI S X.

PA (WANG/) WANG A.

PA (SKEI/) SKEIKY Y A W.

PA (HEPL/) HEPLER W T.
PA (HEND/) HENDERSON R A.
PA (HURA/) HURAL J. P. D.
PA (MCNE/) MCNEILL P. D.
PA (HOUG/) HOUGHTON R L.
PA (DBAS/) Y DE BASSOLS C V.
PA (FOYT/) FOY T M.
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Ranger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA, Hural J;
PI McNeill PD, Houghton RL, Y De BaseolsCV, Foy TM,
XX NPI; 2001-245062/25.
DR Prostate specific protein and its encoding polynucleotide, useful for the
XX treatment and diagnosis of prostate cancer.
PT Example 2; SEQ ID NO 932; 85pp; English.
PS
XX The invention describes a fusion protein comprising at least one amino
CC acid sequence of immunogenic portions of any of the 3 sequences not
CC defined in the specification, or sequences having at least 70 or 90 %
CC sequence identity to any one of the 35 sequences defined in the USPTO web
CC site, which is encoded by any of the 4 nucleotide sequences not defined
CC in the specification. The fusion protein, composition and methods are
CC useful for diagnosing, preventing and/or treating cancer, particularly
CC prostate cancer. The proteins are useful as markers to indicate the
CC presence or absence of cancer. This is the amino acid sequence of a
CC prostate cancer therapy associated protein. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=US20020192763
XX
SQ Sequence 492 AA;

Query Match 99.7%; Score 2709; DB 4; Length 492;

Best Local Similarity 99.8%; Pred. No. 2.8e-202;

Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MALNCGSPPAIGPYENHGYQENPYPAQPTVTVVTVYEVHQAQYPSVPQAPRVLTA 60
Db 1 MALNCGSPPAIGPYENHGYQENPYPAQPTVTVVTVYEVHQAQYPSVPQAPRVLTA 60
Qy 61 SNPVCTQKPSGTVCTSKTKKALCITLTGLTFLVGAALAAAGLLWFKMGSKCSNSGIEC 120
Db 61 SNPVCTQKPSGTVCTSKTKKALCITLTGLTFLVGAALAAAGLLWFKMGSKCSNSGIEC 120
Qy 121 DSSGTCINPNCWCDGVSCHPCGGEDENRCVRLYGNPFILOVYSSQKSWHPVCOQDWNENY 180
Db 121 DSSGTCINPNCWCDGVSCHPCGGEDENRCVRLYGNPFILOVYSSQKSWHPVCOQDWNENY 180
Qy 181 GRAACRDMGYKNNFYSSQGIIVDDSGSTSEWKLNTSAGNVDIYKXLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNNFYSSQGIIVDDSGSTSEWKLNTSAGNVDIYKXLYHSDACSSKAVVSLR 240
Qy 241 CIACGVNLNSRSQSRIVGGESALPGAWPQVSLHVQNVHVCSSGIITPEWIVTAAHCVCK 300
Db 241 CIACGVNLNSRSQSRIVGGESALPGAWPQVSLHVQNVHVCSSGIITPEWIVTAAHCVCK 300
Qy 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVISHPNVDSTKKNNDIALMKLQKELTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVISHPNVDSTKKNNDIALMKLQKELTFNDL 360
Qy 361 VKPVLCPNPGMWLOPEQLCWISGWSGATEBEKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
Db 361 VKPVLCPNPGMWLOPEQLCWISGWSGATEBEKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
Qy 421 TPAMICAGFLQGNVDSQCGSDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVWF 480
Db 421 TPAMICAGFLQGNVDSQCGSDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVWF 480
Qy 481 TDWIYRQMRADG 492

PT diseases, in particular prostate cancer, and as markers for the
PT progression of cancer.
XX
PS Claim 2; SEQ ID NO 932; 87pp; English.
XX
CC The present invention provides prostate-specific coding sequences and
CC their encoded proteins. These can be used in the diagnosis and treatment
CC of cancers, particularly prostate cancer. The present sequence is a
CC protein described in the invention
XX
SQ Sequence 492 AA;
Query Match 99.7%; Score 2709; DB 5; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.8e-202;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVFTVEVHPAQYYPSPVQYAPRVLTQA 60
DB 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVFTVEVHPAQYYPSPVQYAPRVLTQA 60
QY 61 SNPVVCTOPKSPSGTVCSTKTKALCITLTGTLVGAALAGLLKPMGSKCSNGIEC 120
DB 61 SNPVVCTOPKSPSGTVCSTKTKALCITLTGTLVGAALAGLLKPMGSKCSNGIEC 120
QY 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGNFIIQVYSSORCKSHHPVCQDDWNNY 180
DB 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGNFIIQVYSSORCKSHHPVCQDDWNNY 180
QY 181 GSAACRDMGYKNNFYSSQGI VDDSGSTSFMKLNTSAGNVDIYKKLYHSDACSSKAVVSLR 240
DB 181 GSAACRDMGYKNNFYSSQGI VDDSGSTSFMKLNTSAGNVDIYKKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGSGALPGAWPQVSLHVQNVHVGCGSIITPEWITVTAHCVCK 300
DB 241 CIACGVNLNSRQSRIVGSGALPGAWPQVSLHVQNVHVGCGSIITPEWITVTAHCVCK 300
QY 301 PLNNPWHMTAFAGILRQSFMYGAGYQVEKVISHPNYSKTKNNIDIALMKLQKPLTFNDL 360
DB 301 PLNNPWHMTAFAGILRQSFMYGAGYQVEKVISHPNYSKTKNNIDIALMKLQKPLTFNDL 360
QY 361 VKPVCIPNPGMMLQPEQLCWISGWGATEEKGKTSVLNAAKVLLTETORCNSRYVDNLI 420
DB 361 VKPVCIPNPGMMLQPEQLCWISGWGATEEKGKTSVLNAAKVLLTETORCNSRYVDNLI 420
QY 421 TPAMICAGFLQGNVDSQQDGGPLVTSKNNIWLIGDTSWGGCGCAKAYRPGVYGNVMVF 480
DB 421 TPAMICAGFLQGNVDSQQDGGPLVTSKNNIWLIGDTSWGGCGCAKAYRPGVYGNVMVF 480
QY 481 TDWIYRQMRADG 492
DB 481 TDWIYRQMRADG 492
RESULT 14
ABR54532
ID ABR54532 standard; protein; 492 AA.
XX
AC ABR54532;
XX
DT 28-AUG-2003 (first entry)
XX
DE Prostate tumour specific protein SEQ ID 932.
XX Cytostatic; gene therapy; prostate-specific protein; PSP; human;
KW immune response; prostate cancer.
XX
OS Homo sapiens.
XX
PI WO200289747-A2.
XX
PD 14-NOV-2002.
XX
PF 09-MAY-2002; 2002WO-US014753.

Db 481 TDWIYRQMRADG 492
|||||
RESULT 13
ABR95420
ID ABR95420 standard; protein; 492 AA.
XX
AC ABR95420;
XX
DT 19-JUL-2002 (first entry)
XX
DE Human P1000C SEQ ID NO 932.
XX
KW Human; cancer; prostate cancer; vaccine; cytostatic; immunostimulant;
KW gene therapy.
XX
OS Homo sapiens.
XX
FN US2002022248-A1.
XX
PD 21-FEB-2002.
XX
PF 12-JAN-2001; 2001US-00759143.
XX
PR 25-FEB-1997; 97US-00806099.
PR 01-AUG-1997; 97US-00904804.
PR 10-FEB-1998; 98US-00020956.
PR 25-FEB-1998; 98US-00030607.
PR 14-JUL-1998; 98US-00115453.
PR 23-SEP-1998; 98US-00159812.
PR 15-JAN-1999; 99US-00232149.
PR 09-APR-1999; 99US-00288946.
PR 13-JUL-1999; 99US-00352616.
PR 12-NOV-1999; 99US-00439313.
PR 18-NOV-1999; 99US-00443686.
PR 14-JAN-2000; 2000US-00483672.
PR 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 17-JUN-2000; 2000US-00593793.
PR 23-JUN-2000; 2000US-00605783.
PR 10-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00851436.
PR 06-SEP-2000; 2000US-00857279.
PR 02-OCT-2000; 2000US-00879426.
PR 10-OCT-2000; 2000US-00685166.
XX
PA (XUJ/) XU J.
PA (DILL/) DILLON D C.
PA (MITC/) MITCHAM J L.
PA (HARL/) HARLOCKER S L.
PA (JIAN/) JIANG Y.
PA (KALO/) KALOS M D.
PA (FANG/) FANGER G R.
PA (RETT/) RETTER M W.
PA (STOL/) STOLK J A.
PA (DAYC/) DAY C H.
PA (VEDV/) VEDVICK T S.
PA (CART/) CARTER D.
PA (LISX/) LI S X.
PA (WANG/) WANG A.
PA (SKEI/) SKEIKY Y A W.
PA (HEPL/) HEPLER W T.
PA (HEND/) HENDERSON R A.
XX
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
XX WPI; 2002-255649/30.
XX
PT New prostate-specific polynucleotides for diagnosing and treating

XX 09-MAY-2001; 2001US-00852911.
PR 29-JUN-2001; 2001US-00895814.
PR 10-DEC-2001; 2001US-00012896.
XX (CORI-) CORIXA CORP.
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Henderson RA;
PI Kalos MD, Ronger GR, Retter MW, Stolk JA, Day CH, Vedvick IS;
PI Carter D, Li SX, Wang A, Skeiky YAW, Hepler WT, Hural J;
PI McNeill PD, Houghton RL, Vinals Y De BassolsC, Foy TM, Watanabe Y;
PI Deng T;
XX WPI; 2003-167130/16.
XX New prostate-specific proteins and genes, useful in gene therapy,
PT particularly for stimulating an immune response in a patient, or treating
PT prostate cancer in a patient, as well as for diagnosing prostate cancer
PT in a patient.
XX Example 2; Page 643-643; 691pp; English.
XX The present invention relates to novel prostate-specific proteins (PSP)
CC and their coding sequences. The PSPs and their coding sequences are
CC useful for stimulating an immune response in a patient, or for treating
CC prostate cancer in a patient and for determining, detecting or diagnosing
CC the presence of a cancer in a patient. The present sequence was used to
CC illustrate the invention
XX
XX Sequence 492 AA;
SQ Query Match 99.7%; Score 2709; DB 6; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.8e-202;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALNCGSPPAIGPYENHGYQENPYQAPQVTVYVHPAQYPSVPQYAPRVLTQA 60
DB 1 MALNCGSPPAIGPYENHGYQENPYQAPQVTVYVHPAQYPSVPQYAPRVLTQA 60
QY 61 SNPVCTQPKSPSGTCTSKTKALCITLTGTLVGAALAAAGLLKFMGSKCSNGIEC 120
DB 61 SNPVCTQPKSPSGTCTSKTKALCITLTGTLVGAALAAAGLLKFMGSKCSNGIEC 120
QY 121 DSSGTCINPSNWCNCGVSHCPGGEDENCVRLYGSNFIQVYSSQKSWHPVCCDDWNEY 180
DB 121 DSSGTCINPSNWCNCGVSHCPGGEDENCVRLYGSNFIQVYSSQKSWHPVCCDDWNEY 180
QY 181 GRAACRDMGYKNFYSSQGIYVDSGTSFMKLNSTAGNVDIYKLYHSDACSSKAVVSLR 240
DB 181 GRAACRDMGYKNFYSSQGIYVDSGTSFMKLNSTAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGGSALPGAMPQVSLHVQNVHVCVGGSIITPEWIVTAAHCVEK 300
DB 241 CIACGVNLNSRQSRIVGGSALPGAMPQVSLHVQNVHVCVGGSIITPEWIVTAAHCVEK 300
QY 301 PLNPNHWTAPAGTLRSGFMFYGAGYQVEKVI SHPNVDYKTKNDIALMKLQKPLTFNDL 360
DB 301 PLNPNHWTAPAGTLRSGFMFYGAGYQVEKVI SHPNVDYKTKNDIALMKLQKPLTFNDL 360
QY 361 VKPYCLNPNGMQLQPEQLCWISGWGATEEKGKTSSEVLNAAKLLIETQRCNSRYVDNLI 420
DB 361 VKPYCLNPNGMQLQPEQLCWISGWGATEEKGKTSSEVLNAAKLLIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWWLIGDTSWGSACAKAYRPGVGNVWF 480
DB 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWWLIGDTSWGSACAKAYRPGVGNVWF 480
QY 481 TDWIYRQMRADG 492
DB 481 TDWIYRQMRADG 492

RESULT 15

ADB14382
ID ADB14382 standard; protein; 492 AA.
XX
AC ADB14382;
XX
DT 18-DEC-2003 (first entry)
XX Human prostate specific protein P1000C.
XX
XX Human; prostate specific cDNA; cytostatic; immunostimulant; gene therapy;
KW cell therapy; vaccine; T-cell epitope;
KW class I major histocompatibility complex allele; MHC; prostate cancer;
KW tumour; antigen presenting cell.
XX
OS Homo sapiens.
XX
PN US2003185830-A1.
XX
PD 02-OCT-2003.
XX
PF 12-NOV-2002; 2002US-00294025.
XX
XX 25-FEB-1997; 97US-00806099.
PR 01-AUG-1997; 97US-00904804.
PR 09-FEB-1998; 98US-00020956.
PR 25-FEB-1998; 98US-00030607.
PR 14-JUL-1998; 98US-00115453.
PR 23-SEP-1998; 98US-00159812.
PR 15-JAN-1999; 99US-00232149.
PR 09-APR-1999; 99US-00288946.
PR 13-JUL-1999; 99US-00352616.
PR 12-NOV-1999; 99US-00439313.
PR 18-NOV-1999; 99US-00443686.
PR 14-JAN-2000; 2000US-00483672.
PR 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
PR 27-JUN-2000; 2000US-00605783.
PR 09-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
PR 09-NOV-2000; 2000US-00709729.
PR 12-JAN-2001; 2001US-00759143.
PR 09-FEB-2001; 2001US-00780669.
PR 09-MAY-2001; 2001US-00852911.
PR 29-JUN-2001; 2001US-00895814.
PR 10-DEC-2001; 2001US-00012896.
PR 09-MAY-2002; 2002US-00144678.
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Stolk JA, Kalos MD;
PI
XX WPI; 2003-756193/71.
XX N-PSDB; ADB14379, ADB14380, ADB14381.
XX New isolated polypeptide for use in a vaccine for stimulating an immune
PT response, or for treating or diagnosis cancer, preferably prostate
PT cancer.
XX
XX Example 2; Page; 101pp; English.
XX
XX The invention relates to an isolated polypeptide comprising no more than
CC 11-542 amino acids of ADB13563 comprising a sequence ADB14487. The
CC peptides comprise a fragment ADB13563 of that contain naturally processed
CC T-cell epitopes for 3 class I major histocompatibility complex (MHC)
CC alleles. ADB13563 is a polypeptide encoded by a human prostate specific
CC cDNA, one of 648 disclosed as new. Also included are nucleic acids
CC encoding the proteins and peptides, expression vectors, a host cell
CC transformed with the vector, an isolated antibody (or antigen binding

